THE UNIVERSITY OF WARWICK

Feb 2014 No.184

Fuzzy Changes-in Changes

Clément de Chaisemartin (University of Warwick)
and Xavier D'Haultfoeuille (CREST)

WORKING PAPER SERIES

Centre for Competitive Advantage in the Global Economy

Department of Economics

CAGE
Competitive Advantage in the Global Economy

E·S·R·C
ECONOMIC & SOCIAL RESEARCH COUNCIL

Fuzzy Changes-in-Changes*

Clément de Chaisemartin[†] Xavier D'Haultfœuille[‡]

February 4, 2014

Abstract

The changes-in-changes model extends the widely used difference-in-differences to situations where outcomes may evolve heterogeneously. Contrary to difference-in-differences, this model is invariant to the scaling of the outcome. This paper develops an instrumental variable changes-in-changes model, to allow for situations in which perfect control and treatment groups cannot be defined, so that some units may be treated in the "control group", while some units may remain untreated in the "treatment group". This is the case for instance with repeated cross sections, if the treatment is not tied to a strict rule. Under a mild strengthening of the changes-in-changes model, treatment effects in a population of compliers are point identified when the treatment rate does not change in the control group, and partially identified otherwise. Simple plug-in estimators of treatment effects are proposed. We show that they are asymptotically normal, and that the bootstrap is valid. Finally, we use our results to reanalyze findings in Field (2007) and Duflo (2001).

Keywords: differences-in-differences, changes-in-changes, imperfect compliance, instrumental variables, quantile treatment effects, partial identification.

JEL Codes: C21, C23

*We are very grateful to Esther Duflo and Erica Field for sharing their data with us. We also want to thank Alberto Abadie, Joshua Angrist, Stéphane Bonhomme, Marc Gurgand, Guido Imbens, Thierry Magnac, Blaise Melly, Roland Rathelot, Bernard Salanié, Frank Vella, Fabian Waldinger, participants at the 7th IZA Conference on Labor Market Policy Evaluation, North American and European Summer Meetings of the Econometric Society, 11th Doctoral Workshop in Economic Theory and Econometrics and seminar participants at Boston University, Brown University, Columbia University, CREST, MIT, Paris School of Economics and St Gallen University for their helpful comments.

[†]Warwick University, clement.de-chaisemartin@warwick.ac.uk

[‡]CREST, xavier.dhaultfoeuille@ensae.fr

1 Introduction

Difference-in-differences (DID) is one of the most popular methods for evaluating the effect of a treatment in the absence of experimental data. It exploits a temporal change in treatment allocation, for instance following a legislative change. In its basic version, a "control group" is untreated at two dates, whereas a "treatment group" becomes treated at the second date. If the effect of time is the same in both groups, the so-called common trend assumption, one can measure the effect of the treatment on the treated by comparing the evolution of the outcome in both groups. DID only require repeated cross section data, not necessarily panel data, which may explain why this method is so pervasive.

Notwithstanding, the common trend assumption raises a number of concerns. If the control and treatment groups are different and the effect of time is heterogenous, the common trend condition is unlikely to hold. Suppose for instance that one studies the effect of job training on wages, using data where low-wage workers benefit from job training after a given date. If high wages increase more on average than low wages during the period at stake, the common trend assumption fails to hold. Besides, the common trend assumption is not invariant to monotonic transformations of the outcome. As shown by Athey & Imbens (2002), this assumption requires that the effect of time and group on the outcome be additively separable, which cannot be true for both the outcome and its logarithm. This leads to the logs versus levels problem: when considering the level of the outcome or its growth rate, treatment effects estimated through DID may considerably change. For instance, Meyer et al. (1995) find no significant effect of injury benefits on injury duration, while they find strong effects on the logarithm of injury duration.

To deal with this problem, Athey & Imbens (2006) consider a nonlinear extension of difference-in-differences, the changes-in-changes (CIC) model.[1] It relies on the assumption that a control and a treatment unit with the same outcome at the first period would also have had the same outcome at the second period if the treatment unit had then not been treated. Hereafter, we refer to this condition as the common change assumption. This condition allows for heterogeneous effects of time: people with different outcomes at the first period can experience different evolutions over time. And contrary to the common trend assumption, the common change assumption is invariant to monotonic transforms of potential outcomes.

In this paper, we develop a framework that extends the CIC model to fuzzy situations in which the treatment rate increases more in one group than in the other. Many natural experiments

[1]Their estimator is closely related to an estimator proposed by Juhn et al. (1993) and Altonji & Blank (2000) to decompose the Black-White wage differential into changes in the returns to skills and changes in the relative skill distribution.

cannot be analyzed within the standard DID or CIC framework. They do not lead to a sharp change in treatment rate for any group defined by a set of observable characteristics, but only to a larger increase of the treatment rate in some groups than in others. With panel data at hand, the analyst could define the treatment group as units going from non treatment to treatment between the two periods, while the control group could be made up of units remaining untreated at the two periods. But this definition of groups would be endogeneous, and might violate the common trend assumption. Units choosing to go from non treatment to treatment between the two periods might do so because they experience different trends in outcomes.

In such settings, the standard practice is to use linear instrumental variable (IV) regressions to estimate treatment effects. A good example is Duflo (2001). She considers a school construction program in Indonesia which led to the construction of more schools in districts where few schools were previously available. She defines control districts as those in which many schools were already available previous to the program, while treatment districts are those in which few schools were available. Because more schools were constructed in treatment districts, years of schooling increased more in those districts. The author then estimates returns to schooling through an IV regression in which time and group fixed effects are used as included instruments for treatment, while the excluded instrument is the interaction of time and group. The resulting coefficient for treatment in this "IV-DID" regression is the ratio of the DID on the outcome and on treatment, which is sometimes referred to as the Wald-DID. Similarly, Lochner & Moretti (2004) use state compulsory laws as an instrument for schooling. They also estimate IV regressions with time and group fixed effects as included instruments, so their coefficient of interest is a weighted average of Wald-DID across groups and periods of time. Other examples include Burgess & Pande (2005), Field (2007), or Akerman et al. (2013), who estimate similar type of IV regressions as in Lochner & Moretti (2004).

de Chaisemartin (2013) studies the conditions under which these IV-DID regressions capture some treatment effect parameter. He first considers a simple model with constant effect of time in which the effect of the treatment can be heterogeneous across groups but is homogeneous within groups. Identification by IV-DID can fail in this model. Assume for instance that the effect of the treatment is strictly positive in the two groups and twice as large in the control than in the treatment group. Assume also that the treatment rate increased twice as much in the treatment than in the control group. Then, the Wald-DID will be equal to 0: the lower increase of the treatment rate in the control group is exactly compensated by the fact that the effect of the treatment is higher in this group. The Wald-DID does not capture the effect of the treatment in any of the two groups, or a weighted average of the two. Therefore, in this simple model, identification by IV-DID requires that the effect of the treatment be the same across groups. He then shows that this result carries to more general models with heterogeneous effects. In those models, identification by IV-DID requires standard common trend assumptions, but it also requires that the average effect of the treatment be the same in the treatment and in the control groups, at least among observations who switch treatment status over time.

To circumvent these shortcomings, we study an instrumental variable changes-in-changes (IV-CIC) model which does not require common trend assumptions, is invariant to monotonic transforms of the outcome, and does not impose that some subgroups of observations in the treatment and in the control groups have the same treatment effects. Our model combines both an increasing production function for the outcome, as in Athey & Imbens (2006), and a latent index model for treatment choice in the spirit of Vytlacil (2002). Relative to Athey & Imbens (2006), the main supplementary ingredient we impose is a strengthening of the common change assumption. Formally, we impose that both potential outcomes and the propensity to be treated satisfy the common change assumption. Importantly, this allows for endogenous selection, including Roy models where potential outcomes evolve heterogeneously.

In this framework, we show that the marginal distributions of potential outcomes for compliers are point identified if the treatment rate remains constant in the control group, and partially identified otherwise. The intuition for this result goes as follows. When the treatment rate is constant in the control group, any change in the distribution of the outcome of this group can be attributed to time. By the common change assumption, time has the same effect in both groups among individuals with the same outcome. We can therefore use the control group to identify the effect of time, and remove this effect in the treatment group. Any remaining change in the distribution of the outcome in the treatment group can then be attributed to the increase in treatment rate it experienced over time. Thus, the marginal distributions of potential outcomes for compliers are identified. But when the treatment rate is not constant in the control group, the evolution of the outcome in this group may stem both from the effect of time and from the change in the treatment rate. Therefore, the effect of time is only partially identified, which in turn implies that the marginal distributions of potential outcomes for compliers are partially identified as well. We exhibit bounds on these distributions, and show that they are sharp under testable monotonicity conditions. The smaller the change of the treatment rate in the control group, the tighter the bounds.

We then consider three extensions of those main results. Firstly, we show how to incorporate covariates in the analysis. Secondly, we show how to extend our analysis to settings with many time periods and many groups. Having many groups might allow recovering point identification: the more groups there are, the more likely it is that the treatment rate does not change in at least one of them. Thirdly, we show that our IV-CIC model is testable. When it is point identified, its testable implication is very similar to the testable implication of the IV model with binary treatment and instrument in Angrist et al. (1996), which has been studied by Kitagawa (2013). But when the model is partially identified, its testable implication takes a different form.

We also develop inference on average and quantile treatment effects. Using the functional delta method, we show that simple plug-in estimators of treatment effects in the fully identified case, and of the bounds in the partially identified one, are asymptotically normal under mild conditions. Because the variance takes a complicated form, the bootstrap is convenient to use here, and we prove that it is consistent.

Finally, we apply our results to two different data sets. We first revisit Field (2007), who studies the effect of granting property titles to urban squatters on their labor supply. As the treatment rate is stable in the comparison group used by the author, we are in the point identified case. Our IV-CIC model allows us to study distributional effects of the treatment which were not studied by the author. We show that property rights have a stronger relative effect on households with a low initial labor supply. A possible explanation is that among squatters, only one household member has to stay home to look after the household's residence, irrespective of the household size. The effect of the program would then be large for small households with low initial labor supply, and smaller otherwise. Knowing this pattern of heterogeneity might have substantial consequences on social choice. A utilitarian social planner will indeed be more prone to implementing a titling program with heterogeneous than with constant relative effects, provided utility of agents is concave in individual consumption. We then revisit results in Duflo (2001) on returns to education. As the treatment rate changes in the comparison group used by the author, we are in the partially identified case. Our bounds are wide and uninformative, because the treatment rate increased substantially in the control group. Our IV-CIC model does not allow us to draw informative conclusions on returns to education from this natural experiment.

Our paper therefore shows that in fuzzy settings, researchers must find a control group in which the treatment rate is stable over time to point identify treatment effects under our non linear IV-CIC model. This will be possible to achieve when a group is excluded from treatment at both dates. This might also be possible to achieve when a policy is extended to a previously ineligible group, or when a program or a technology previously available in some geographic areas is extended to others (see e.g. Field, 2007, or Akerman et al., 2013). When the treatment rate slightly changes in the control group, researchers can still derive informative bounds for treatment effects, as we show in a third application developed in Appendix C. When exposition to treatment substantially changes in the control group as well, using our IV-CIC model will result in wide and uninformative bounds. In such instances, point identification can still be achieved using IV-DID, but at the price of imposing more stringent conditions.

Besides Athey & Imbens (2006), our paper is related to D'Haultfoeuille et al. (2013), who study the possibly nonlinear effects of a continuous treatment using repeated cross sections. They also rely on a control group to identify the effect of time, but in their case the choice of this group is driven by the data. Our paper is also connected to several recent papers analyzing difference-in-differences models. de Chaisemartin (2013) studies the identifying assumptions underlying IV-DID regressions. Several recent papers have also considered different routes from the one taken in Athey & Imbens (2006) to weaken the common trend condition in "sharp" DID. Blundell et al. (2004) and Abadie (2005) consider a conditional version of this assumption, and adjust for covariates using propensity score methods. Donald & Lang (2007) and Manski & Pepper (2012) allow for some variations in the way time affects the control and treatment groups, provided these variations satisfy some restrictions. Bonhomme & Sauder (2011) consider a linear model allowing for heterogeneous effects of time, and show how it can be identified using an instrument.

The remainder of the paper is organized as follows. Section 2 presents our model. Section 3 is devoted to identification. Section 4 presents some extensions. Section 5 deals with inference. In section 6 we apply our results to the two aforementioned data sets. Section 7 concludes. The appendix gathers all the proofs, some technical lemmas and a third application on the effect of a pharmacotherapy on smoking cessation.

2 The instrumental variable Changes-in-Changes model

Let $T \in \{0;1\}$ denote time and G denote the dummy of the treatment group (so that $G = 0$ for the control group). The treatment D is supposed to be binary. Hereafter, for any random variables R and S, $R \sim S$ means that R and S have the same probability distribution. $\mathcal{S}(R)$ and $\mathcal{S}(R|S)$ denote respectively the support of R and the support of R conditional on S. As Athey & Imbens (2006), we introduce for any random variable R the corresponding random variables R_{gt} such that

$$R_{gt} \sim R|\, G = g, T = t.$$

Let F_R and $F_{R|S}$ denote the cumulative distribution function (cdf) of R and its cdf conditional on S. For any event A, $F_{R|A}$ is the cdf of R conditional on A. With a slight abuse of notation, $P(A)F_{R|A}$ should be understood as 0 when $P(A) = 0$. For any increasing function F on the real line, we denote by F^{-1} its generalized inverse:

$$F^{-1}(q) = \inf \{x \in \mathbb{R} / F(x) \geq q\}.$$

In particular, F_X^{-1} is the quantile function of X. We adopt the convention that $F_X^{-1}(q) = \inf \mathcal{S}(X)$ for $q < 0$, and $F_X^{-1}(q) = \sup \mathcal{S}(X)$ for $q > 1$.

In our IV-CIC model, $D \neq G \times T$ in general. Some units may be treated in the control group or at period 0, and all units are not necessarily treated in the treatment group at period 1. This will arise when repeated cross sections are available, and the treatment is not tied to a strict rule. In such instances, it is not possible to know whether units at the second period were treated in the first period, and we cannot define a control group that was completely untreated in the first period. When panel data are available, it may not be desirable to define the treatment group as units going from non treatment to treatment, and the control group as units untreated at both periods. With this definition of groups, the assumptions underlying both the changes-in-changes model and the difference-in-differences model will be violated if individuals become treated because of an Ashenfelter's dip. Defining the control and treatment groups in an exogenous way usually makes the identifying assumption more credible. We let $\lambda_d = P(D_{01} = d)/P(D_{00} = d)$ be the ratio of the shares of people receiving treatment d in period 1 and period 0 in the control group. For instance, $\lambda_0 > 1$ when the share of untreated observations increases in the control group between period 0 and 1. $\lambda_0 > 1$ implies that $\lambda_1 < 1$ and conversely. $\mu_d = P(D_{11} = d)/P(D_{10} = d)$ is the equivalent of λ_d for the treatment group.

We assume that at period 1, individuals from the treatment group receive extra incentives to get treated. We model this by introducing the binary instrument $Z = T \times G$. In Duflo (2001), Z is an intensive school construction program which was implemented in treatment districts in period 1 and not in control ones, giving greater incentives to go to school to children living in those districts. The two corresponding potential treatments, $D(1)$ and $D(0)$, stand for the treatment an individual would choose to receive with and without this supplementary incentive. The observed treatment is $D = ZD(1) + (1 - Z)D(0)$. $Y(1)$ and $Y(0)$ are the potential outcomes of an individual with and without treatment. Implicit in this notation is the exclusion restriction that the instrument does not affect the outcome directly. The observed outcome is $Y = DY(1) + (1 - D)Y(0)$. As in Athey & Imbens (2006), we consider the following model for the potential outcomes:

$$Y(d) = h_d(U_d, T), \quad d \in \{0; 1\}. \tag{1}$$

We also make the following restrictions.

Assumption 1 *(Monotonicity)*

$h_d(u,t)$ *is strictly increasing in* u *for all* $(d,t) \in \{0,1\}^2$.

Assumption 2 *(Latent index model for potential treatments)*

$D(z) = 1\{V \geq v_z(T)\}$ *with* $v_0(t) > v_1(t)$ *for* $t \in \{0;1\}$.

Assumption 3 *(Time invariance within groups)*

For $d \in \{0,1\}$, $(U_d, V) \perp\!\!\!\perp T \,|\, G$.

Remarks on these assumptions are in order. Under Assumptions 1 and 2, V can be interpreted as a propensity for treatment. Similarly, if potential outcomes were schooling performances or wages, U_d could be interpreted as an ability index, and we stick to this interpretation hereafter. Our latent index model is the same as in Vytlacil (2002), except that the threshold can depend on time, to allow for the treatment rate to evolve over time. As shown by Vytlacil (2002), such a latent index model is equivalent to the no defiers condition in Imbens & Angrist (1994). Note that our results would not change if U_d and V were indexed by time, except that we would have to rewrite Assumption 3 as follows: for $d \in \{0,1\}$, $\left(U_d^0, V^0\right) |G \sim \left(U_d^1, V^1\right) |G$. This means we could allow individual ability and taste for treatment to change over time, provided their distribution remains the same in each group. Assumption 2 might seem to imply that time can affect individual treatment choice in only one direction. The previous discussion shows that this is actually not necessary for our results to hold. Time might induce some observations to go from non-treatment to treatment, while having the opposite effect on other observations. In what follows, we do not index U_d and V by time to alleviate the notational burden, but it is worth bearing in mind that this is just an expositional choice, not a substantive restriction.

Assumption 3 requires that the joint distribution of ability and propensity for treatment remains stable in each group over time. It implies $U_d \perp\!\!\!\perp T \,|\, G$ and $V \perp\!\!\!\perp T \,|\, G$, which correspond to the time invariance assumption in Athey & Imbens (2006). As a result, Assumptions 1-3 impose a standard CIC model both on Y and D. But Assumption 3 also implies $U_d \perp\!\!\!\perp T \,|\, G, V$, which means that in each group, the distribution of ability among people with a given taste for treatment should not change over time. This is the key supplementary ingredient with respect to the standard CIC model that we are going to use for identification.

This joint independence assumption allows for endogenous selection into treatment, but it restricts the way time can affect the selection mechanism. To understand this better, assume potential treatments follow a pure Roy model: $D(z) = \mathbb{1}\{Y(1) - Y(0) \geq c(z)\}$. Suppose also that $Y(d) = U_d + \eta_d T + \gamma U_d T$, and that the standard CIC assumption is verified:

$$(U_0, U_1) \perp\!\!\!\perp T|G. \tag{2}$$

This model allows for different trends in potential outcomes across ability levels and therefore across groups, because it does not impose that the distribution of U_0 and U_1 be the same in the two groups. It also allows for differential trends for $Y(0)$ and $Y(1)$ through η_0 and η_1. Combining this with the Roy model implies that selection into treatment can change over time. Finally, it satisfies Assumption 1 provided $\gamma > -1$. One can then rewrite

$$D(z) = \mathbb{1}\left\{U_1 - U_0 \geq \frac{c(z) - (\eta_1 - \eta_0)T}{1 + \gamma T}\right\}.$$

Assumption 2 is satisfied with $V = U_1 - U_0$ and $v_z(T) = [c(z) - (\eta_1 - \eta_0)T]/(1 + \gamma T)$. Therefore, $(U_0, U_1) \perp\!\!\!\perp T|G$ implies that Assumption 3 is satisfied, because V is a deterministic function of U_0 and U_1. On the contrary, with γ_d instead of γ in the potential outcomes equation, Assumption 3 cannot hold, because $V = U_1 - U_0 + T(\gamma_1 U_1 - \gamma_0 U_0)$. Assumption 3 is compatible with a Roy model in which time can have heterogeneous effects on outcomes across ability levels, and an homogeneous effect on propensity for treatment. But it is not compatible with a Roy model in which this second effect is heterogeneous across ability levels.

Finally, it is worth mentioning that the IV-DID model is incompatible with the Roy selection model and outcome equations outlined above. As shown in de Chaisemartin (2013), in a model allowing for heterogeneous treatment effects, the IV-DID method relies on common trend assumptions both on potential outcomes and treatments. Common trend on potential treatments is necessary to ensure that the denominator of the Wald-DID ratio captures the size of the population induced to switch from non treatment to treatment between period 0 and 1 because of the instrument, not because of the effect of time alone. In the Roy model above, the common trend on potential outcomes implies that $\gamma = 0$. But even in this case, the common trend on potential treatments does not hold in general. To see this, note that under (2), this condition is equivalent to

$$\begin{aligned} & P(U_1 - U_0 \geq c(z) - (\eta_1 - \eta_0)|G = 1) - P(U_1 - U_0 \geq c(z)|G = 1) \\ = \quad & P(U_1 - U_0 \geq c(z) - (\eta_1 - \eta_0)|G = 0) - P(U_1 - U_0 \geq c(z)|G = 0). \end{aligned}$$

This is unlikely to be true, unless we are ready to assume that $U_1 - U_0 \perp\!\!\!\perp G$. But this would amount to assuming that groups are as good as randomly assigned, in which case we do not need to resort to a longitudinal analysis to capture treatment effects. We could merely use a standard cross-sectional IV using group as an instrument for treatment in period 1.

Hereafter, we refer to Assumptions 1-3 as to the IV-CIC model. This model extends the CIC analysis to the fuzzy case. Actually, one can show that our IV-CIC assumptions reduce to those in Athey & Imbens (2006) when $P(D_{11} = 1) = 1$ and $P(D_{10} = 1) = P(D_{01} = 1) = P(D_{00} = 1) = 0$, namely when their "sharp" setting holds.

Finally, we impose the two following restrictions, which are directly testable in the data.

Assumption 4 *(Data restrictions)*

1. $\mathcal{S}(Y_{gt}|D = d) = \mathcal{S}(Y) = [\underline{y}, \overline{y}]$ *with* $(\underline{y}, \overline{y}) \in \overline{\mathbb{R}}^2$*, for* $(g, t, d) \in \{0; 1\}^3$.

2. $F_{Y_{gt}|D=d}$ *is strictly increasing and continuous on* $\mathcal{S}(Y)$*, for* $(g, t, d) \in \{0; 1\}^3$.

Assumption 5 *(Rank conditions)*

1. $P(D_{11} = 1) - P(D_{10} = 1) > 0$.

2. *If* $P(D_{00} = 0) > 0$, $F_{Y_{10}|D=0} \circ F^{-1}_{Y_{00}|D=0}(\lambda_0) > \mu_0$ *and* $F_{Y_{10}|D=0} \circ F^{-1}_{Y_{00}|D=0}(1-\lambda_0) < 1-\mu_0$.

The first condition of Assumption 4 is a common support condition. Athey & Imbens (2006) take a similar assumption and show how to derive partial identification results when it is not verified. Point 2 is satisfied if the distribution of Y is continuous with positive density in each of the eight groups $\times$ period $\times$ treatment status cells.

Assumption 5 corresponds to a rank condition in a standard IV model. The IV-CIC identification strategy requires that treatment rate changes in at least one group. If it diminishes in the two groups over time we can just switch labels and consider $1 - D$ as the treatment variable. Therefore, the first condition is without loss of generality: it does not impose anything except that treatment rate changes in at least one group.

To better understand the second condition of Assumption 5, let us draw a parallel with the IV-DID model. The rank condition in an IV-DID regression is

$$P(D_{11} = 1) - P(D_{10} = 1) - (P(D_{01} = 1) - P(D_{00} = 1)) \neq 0,$$

meaning that the treatment rate does not follow parallel trends in the two groups. If it is not satisfied, one cannot run an IV-DID analysis because the instrument has no effect on treatment. Point 2 in Assumption 5 is similar in spirit to this rank condition. Under Point 1, $\mu_0 < 1$. Therefore, if $\lambda_0 > 1$, Point 2 will automatically hold. $\mu_0 < 1$ and $\lambda_0 > 1$ corresponds to a situation where the share of untreated people decreases in the treatment group while it increases in the control group. Giving the instrument to the treatment group in period 1 has a strong effect on their propensity for treatment as it leads to opposite trends in treatment rates across the two groups. If $\lambda_0 < 1$, this condition will not automatically hold. The share of untreated people decreases in both groups, meaning that the instrument might not have a very large effect on the share of people receiving treatment. If $F_{Y_{10}|D=0} \circ F^{-1}_{Y_{00}|D=0}$ is the identity function, this condition will be verified if and only if $\lambda_0 > \mu_0$. In practice, $F_{Y_{10}|D=0} \circ F^{-1}_{Y_{00}|D=0}$ might not be the identity function, but this still shows that the larger the difference between λ_0 and μ_0, the more likely this condition holds.

Finally, it should be emphasized that as the first condition, the second condition of Assumption 5 is testable from the data, so it can be assessed beforehand by researchers willing to use our IV-CIC model. When this test is rejected, our model cannot be used as it will only yield trivial bounds for treatment effects, as we will explain below.

Before getting to the identification results, it is useful to define five subpopulations of interest. Assumption 3 implies that $P(D_{10} = 1) = P(V \geq v_0(0)|G = 1)$, and similarly $P(D_{11} = 1) = P(V \geq v_1(1)|G = 1)$. Therefore, under Assumption 5, $v_0(0) > v_1(1)$. Similarly, if the treatment rate increases (resp. decreases) in the control group, $v_0(0) > v_0(1)$ (resp. $v_0(0) < v_0(1)$). Finally, assumption Assumption 2 implies $v_1(1) \leq v_0(1)$. Let always takers be such that $V \geq v_0(0)$, and let never takers be such that $V < v_1(1)$. Always takers are units who get treated in period 0 even without receiving any incentive for treatment. Never takers are units who do not get treated in period 1 even after receiving an incentive for treatment. Let $TC = V \in [\min(v_0(0), v_0(1)), \max(v_0(0), v_0(1)))$. TC stands for "time compliers," and represents observations whose treatment status switches between the two periods because of the effect of time. Similarly, let $IC = V \in [v_1(1), v_0(1))$.[2] IC stands for instrument compliers. This population corresponds to compliers as per the definition of Imbens & Angrist (1994), that is to say observations that become treated through the effect of Z only. However, in our IV-CIC model, we cannot learn anything on this population. Instead, our identification results focus on observations that satisfy $V \in [v_1(1), v_0(0))$. This corresponds to untreated observations at period 0 who become treated at period 1, through both the effect of Z and time. We refer to those observations as compliers to simplify the exposition, and we let hereafter C denote the event $V \in [v_1(1), v_0(0))$. If the treatment rate increases in the control group (i.e. if $v_0(1) < v_0(0)$), we merely have $C = IC \cup TC$, while if it decreases we have $C = IC \setminus TC$.

Our parameters of interest are the cdf of $Y(1)$ and $Y(0)$ among compliers, as well as the Local Average Treatment Effet (LATE) and Quantile Treatment Effects (QTE) within this population, which are respectively defined by

$$\begin{aligned} \Delta &= E\left(Y_{11}(1) - Y_{11}(0)|C\right), \\ \tau_q &= F^{-1}_{Y_{11}(1)|C}(q) - F^{-1}_{Y_{11}(0)|C}(q), \quad q \in (0,1). \end{aligned}$$

[2]IC is defined to be empty when $v_0(1) = v_1(1)$.

3 Identification

3.1 Point identification results

We first show that when the treatment rate does not change between the two periods, the cdf of $Y(1)$ and $Y(0)$ among compliers are identified. Consequently, the LATE and QTE are also point identified. Let $Q_d(y) = F^{-1}_{Y_{01}|D=d} \circ F_{Y_{00}|D=d}(y)$ be the the quantile-quantile transform of Y from period 0 to 1 in the control group conditional on $D = d$. This transform maps y at rank q in period 0 into the corresponding y' at rank q as well in period 1. Also, let $Q_D = DQ_1 + (1 - D)Q_0$. Finally, let $H_d(q) = F_{Y_{10}|D=d} \circ F^{-1}_{Y_{00}|D=d}(q)$ be the inverse quantile-quantile transform of Y from the control to the treatment group in period 0 conditional on $D = d$. This transform maps rank q in the control group into the corresponding rank q' in the treatment group with the same value of y.

Theorem 3.1 *If Assumptions 1-5 hold and for $d \in \{0, 1\}$ $P(D_{00} = d) = P(D_{01} = d) > 0$, $F_{Y_{11}(d)|C}(y)$ is identified by*

$$
\begin{aligned}
F_{Y_{11}(d)|C}(y) &= \frac{P(D_{10} = d)F_{Q_d(Y_{10})|D=d}(y) - P(D_{11} = d)F_{Y_{11}|D=d}(y)}{P(D_{10} = d) - P(D_{11} = d)} \\
&= \frac{P(D_{10} = d)H_d \circ F_{Y_{01}|D=d}(y) - P(D_{11} = d)F_{Y_{11}|D=d}(y)}{P(D_{10} = d) - P(D_{11} = d)}.
\end{aligned}
$$

This implies that Δ and τ_q are also identified. Moreover,

$$\Delta = \frac{E(Y_{11}) - E(Q_D(Y_{10}))}{E(D_{11}) - E(D_{10})}.$$

This theorem combines ideas from Imbens & Rubin (1997) and Athey & Imbens (2006). We seek to recover the distribution of $Y(1)$ and $Y(0)$ among compliers in the treatment $\times$ period 1 cell. When the treatment rate does not change in the control group, $v_0(0) = v_0(1)$. As a result, there are no time compliers, and compliers are merely instrument compliers. To recover the distribution of $Y(1)$ among them, we start from the distribution of Y among all treated observations of this cell. As shown in Table 1, those include both compliers and always takers. Consequently, we must "withdraw" from this distribution the cdf of $Y(1)$ among always takers, exactly as in Imbens & Rubin (1997). But this last distribution is not observed. To reconstruct it, we adapt the ideas in Athey & Imbens (2006). As is shown in Table 1, all treated observations in the control group or in period 0 are always takers; the distribution of $Y(1)$ among always takers is identified within those three cells. Since Assumption 3 implies

$$U_1 \perp\!\!\!\perp T|G, V \geq v_0(0),$$

the distribution of U_1 is the same in periods 0 and 1 among always takers in the control group. Similarly, this equation implies that U_1 also has the same distribution in period 0 and 1 among always takers of the treatment group. This implies that the quantile-quantile transform among always takers is the same in the treatment and control groups. As a result, we can identify the distribution of $Y(1)$ among treatment $\times$ period 1 always takers, applying the quantile-quantile transform from period 0 to 1 among treated observations in the control group to the distribution of $Y(1)$ among always takers in the treatment group in period 0. Identification of the distribution of $Y(0)$ among compliers in the treatment $\times$ period 1 cell is obtained through similar steps.

<table>
<tr><th></th><th>Period 0</th><th>Period 1</th></tr>
<tr><td rowspan="2">Control Group</td><td>30% treated: Always Takers</td><td>30% treated: Always Takers</td></tr>
<tr><td>70% untreated: Never Takers and Compliers</td><td>70% untreated: Never Takers and Compliers</td></tr>
<tr><td rowspan="2">Treatment Group</td><td>20% treated: Always Takers</td><td>65% treated: Always Takers and Compliers</td></tr>
<tr><td>80% untreated: Never Takers and Compliers</td><td>35% Untreated: Never Takers</td></tr>
</table>

Table 1: Populations of interest when $P(D_{00} = 0) = P(D_{01} = 0)$.

Another way to understand the transform we use to reconstruct the cdf of $Y(1)$ among always takers is to regard it as a double matching. Consider an always taker in the treatment $\times$ period 0 cell. She is first matched to an always taker in the control $\times$ period 0 cell with same y. Those two always takers are observed at the same period of time and have the same treatment status. Therefore, under assumption Assumption 1 they must have the same u_1. Second, the control $\times$ period 0 always taker is matched to its rank counterpart among always takers of the control $\times$ period 1 cell (this is merely the quantile-quantile transform). We denote y^* the outcome of this last observation. Because $U_1 \perp\!\!\!\perp T|G, V \geq v_0(0)$, those two observations must also have the same u_1. Consequently, $y^* = h_1(u_1, 1)$, which means that y^* is the outcome that the treatment $\times$ period 0 cell always taker would have obtained in period 1. Therefore, to recover the whole distribution of $Y(1)$ in period 1 among test group always takers, we translate the whole distribution of always takers in the period 0 $\times$ test group cell from y to the corresponding y^* for each value of y.

Note that our LATE estimand is similar to the LATE estimand in Imbens & Angrist (1994), the standard Wald ratio. Once noted that conditional on $G = 1$, $Z = T$, we have

$$\Delta = \frac{E(Y|G = 1, Z = 1) - E(Q_D(Y)|G = 1, Z = 0)}{E(D|G = 1, Z = 1) - E(D|G = 1, Z = 0)}.$$

The Wald ratio has the same expression, except that here Y is replaced by $Q_D(Y)$ in the second term of the numerator. The standard Wald parameter does not identify a causal effect here because conditional on $G = 1$, Z (i.e. T) is not independent of $Y(d)$: the distributions of potential outcomes might evolve with time. To take into account the effect of time on the distribution of potential outcomes, we apply the quantile-quantile transform observed between period 0 and 1 in the control group to the distribution of Y in period 0 in the treatment group. Assumptions 1 and 3 ensure that quantile-quantile transforms are the same in the two groups. Likewise, the formulae of the cdf of $Y(1)$ and $Y(0)$ among compliers are very similar to those obtained in Imbens & Rubin (1997). For instance, the cdf of $Y(1)$ rewrites as

$$\frac{P(D = 1|G = 1, Z = 1)F_{Y|D=1,G=1,Z=1}(y) - P(D = 1|G = 1, Z = 0)F_{Q_1(Y)|D=1,G=1,Z=0}(y)}{P(D = 1|G = 1, Z = 1) - P(D = 1|G = 1, Z = 0)}.$$

The cdf of $Y(1)$ in the Imbens and Angrist IV model has the same expression except that $Q_1(Y)$ is replaced by Y in the second term of the numerator. Here again, this is to account for the fact that conditional on $G = 1$, the instrument T is not independent of potential outcomes.

Under Assumptions 1-5, the LATE and QTE for compliers are point identified when $0 < P(D_{00} = 0) = P(D_{01} = 0) < 1$, but not in the extreme cases where $P(D_{00} = 0) = P(D_{01} = 0) \in \{0, 1\}$. For instance, when $P(D_{00} = 1) = P(D_{01} = 1) = 1$, $F_{Y_{11}(1)|C}$ is identified by Theorem 3.1, but $F_{Y_{11}(0)|C}$ is not. Such situations are likely to arise in practice, for instance when a policy is extended to a previously ineligible group, or when a program or a technology previously available in some geographic areas is extended to others (see Subsection 6.1 below). We therefore consider a mild strengthening of our assumptions under which both $F_{Y_{11}(0)|C}$ and $F_{Y_{11}(1)|C}$ are point identified in those instances.

Assumption 6 *(Common effect of time on both potential outcomes)* $h_0(u, t) = h_1(u, t) = h(u, t)$ *for every* $(u, t) \in \mathcal{S}(U) \times \{0, 1\}$.

Assumption 6 requires that the effect of time be the same on both potential outcomes. It implies that two observations with the same outcome in period 0 will also have the same outcome in period 1 if they do not switch treatment between the two periods, even if they do not share the same treatment at period 0. Under this assumption, if $P(D_{00} = 1) = P(D_{01} = 1) = 1$, changes in the distribution of Y in the control group over time allow us to identify the effect of time both on $Y(0)$ and $Y(1)$, hence allowing us to recover both $F_{Y_{11}(0)|C}$ and $F_{Y_{11}(1)|C}$.

Theorem 3.2 *If Assumptions 1-6 hold and $P(D_{00} = d) = P(D_{01} = d) = 0$ for some $d \in \{0,1\}$, $F_{Y_{11}(d)|C}(y)$ and $F_{Y_{11}(1-d)|C}(y)$ are identified by*

$$F_{Y_{11}(d)|C}(y) = \frac{P(D_{10} = d)F_{Q_{1-d}(Y)_{10}|D=d}(y) - P(D_{11} = d)F_{Y_{11}|D=d}(y)}{P(D_{10} = d) - P(D_{11} = d)}$$
$$F_{Y_{11}(1-d)|C}(y) = \frac{P(D_{10} = 1-d)F_{Q_{1-d}(Y)_{10}|D=1-d}(y) - P(D_{11} = 1-d)F_{Y_{11}|D=1-d}(y)}{P(D_{10} = 1-d) - P(D_{11} = 1-d)}.$$

This implies that Δ and τ_q are also identified. Moreover,

$$\Delta = \frac{E(Y_{11}) - E(Q_{1-d}(Y_{10}))}{E(D_{11}) - E(D_{10})}.$$

A last situation worth noting is when the treatment rate is equal to 0 at both dates in the control group, and is also equal to 0 in the first period in the treatment group. This is a special case of Theorem 3.2, but in such instances we can actually identify the model under fewer assumptions. To see this, note that in such situations,

$$F_{Y_{11}(1)|C} = F_{Y_{11}|D=1} \tag{3}$$

because there are no always takers in the treatment group. Therefore, we only need to recover $F_{Y_{11}(0)|C}$. But since the distribution of $Y_{11}(0)$ among never takers is identified by $F_{Y_{11}|D=0}$, under Assumption 2 we only need to recover $F_{Y_{11}(0)}$. This can be achieved under the standard changes-in-changes assumptions, as the control group remains fully untreated at both dates.

3.2 Partial identification

When $P(D_{00} = d) = P(D_{01} = d)$, $F_{Y_{11}(d)|C}$ is identified under Assumptions 1-5 or Assumptions 1-6. We shall show below that if this condition is not verified, the functions $F_{Y_{11}(d)|C}$ are partially identified. For that purpose, we must distinguish between two cases.

The first one is when $P(D_{00} = d) > 0$. In such instances, the first of the two matchings described in the previous section works as before. But the second one collapses, since we no longer have $v_0(1) = v_0(0)$. Among treated observations in the control $\times$ period 0 cell, U_1 is distributed conditional on $G = 0, V \geq v_0(0)$, while it is distributed conditional on $G = 0, V \geq v_0(1)$ in period 1. This implies that we cannot match period 0 and period 1 observations on their rank. For instance, when the treatment rate increases in the control group, treated observations in the control group include only always takers in period 0, while in period 1 they also include time compliers, as is shown in Table 2. However, under Assumption 3 the share of time compliers among treated observations in the control group in period 1 is known. Therefore, under Assumptions 1-5, the distributions of potential outcomes among compliers can be written as functions of observed distributions and of $F_{Y_{01}(d)|TC}$, in a formula where $F_{Y_{01}(d)|TC}$ enters with a weight identified from the data.

	Period 0	Period 1
Control Group	30% treated: Always Takers	35% treated: Always Takers and Time Compliers
	70% untreated: Never Takers, Instrument Compliers and Time Compliers	65% untreated: Never Takers and Instrument Compliers
Treatment Group	25% treated: Always Takers	60% treated: Always Takers, Instrument Compliers and Time Compliers
	75% untreated: Never Takers, Instrument Compliers and Time Compliers	40% Untreated: Never Takers

$\mathrm{P}(\mathrm{D}_{01} = 1) \geq \mathrm{P}(\mathrm{D}_{00} = 1)$

	Period 0	Period 1
Control Group	35% treated: Always Takers and Time Compliers	30% treated: Always Takers
	65% untreated: Never Takers and Instrument Compliers	70% untreated: Never Takers, Instrument Compliers and Time Compliers
Treatment Group	25% treated: Always Takers and Time Compliers	60% treated: Always Takers and Instrument Compliers
	75% untreated: Never Takers and Instrument Compliers.	40% Untreated: Never Takers and Time Compliers

$\mathrm{P}(\mathrm{D}_{01} = 1) < \mathrm{P}(\mathrm{D}_{00} = 1)$

Table 2: Populations of interest.

The second case we have to consider is when $P(D_{00} = d) = 0$. In this case, the first step of the aforementioned double matching collapses for the distribution of $Y(d)$. For instance, if $P(D_{00} = 1) = 0$, there are no treated observations in the control group in period 0 to which treated observations in the treatment group in period 0 can be matched. Still, the cdf of Y among treated observations in the treatment $\times$ period 1 cell writes as a weighted average of the cdf of $Y(d)$ among compliers and always or never takers. We can use this fact to bound $F_{Y_{11}(d)|C}$.

The following lemma summarizes these results. To derive bounds on $F_{Y_{11}(d)|C}$ and then on the LATE and QTE, we first relate these cdf to observed distributions and one unidentified cdf.

Lemma 3.1 *If Assumptions 1-5 hold, then:*

- *If* $P(D_{00}=d)>0$,

$$F_{Y_{11}(d)|C}(y)=\frac{P(D_{10}=d)H_d\circ(\lambda_d F_{Y_{01}|D=d}(y)+(1-\lambda_d)F_{Y_{01}(d)|TC}(y))-P(D_{11}=d)F_{Y_{11}|D=d}(y)}{P(D_{10}=d)-P(D_{11}=d)}.$$

- *If* $P(D_{00}=d)=0$,

$$F_{Y_{11}(d)|C}=\frac{P(D_{10}=d)F_{Y_{11}(d)|(2d-1)V>(2d-1)v_0(0)}-P(D_{11}=d)F_{Y_{11}|D=d}}{P(D_{10}=d)-P(D_{11}=d)}.$$

From this lemma, it appears that when $P(D_{00}=d)>0$, we merely need to bound $F_{Y_{01}(d)|TC}$ to derive bounds on $F_{Y_{11}(d)|C}$. In order to do so, we must take into account the fact that $F_{Y_{01}(d)|TC}$ is related to two other cdf. To alleviate the notational burden, let $T_d=F_{Y_{01}(d)|TC}$, $C_d(T_d)=F_{Y_{11}(d)|C}$, $G_0(T_0)=F_{Y_{01}(0)|V<v_0(0)}$ and $G_1(T_1)=F_{Y_{01}(1)|V\geq v_0(0)}$. With those notations, we have

$$\begin{aligned}G_d(T_d) &= \lambda_d F_{Y_{01}|D=d}+(1-\lambda_d)T_d\\ C_d(T_d) &= \frac{P(D_{10}=d)H_d\circ G_d(T_d)-P(D_{11}=d)F_{Y_{11}|D=d}}{P(D_{10}=d)-P(D_{11}=d)}.\end{aligned}$$

The fact that T_d, $G_d(T_d)$ and $C_d(T_d)$ should all be included between 0 and 1 imposes several restrictions on T_d, from which we derive our bounds. Let $M_0(x)=\max(0,x)$, $m_1(x)=\min(1,x)$ and define

$$\begin{aligned}\underline{T}_d &= M_0\left(m_1\left(\frac{\lambda_d F_{Y_{01}|D=d}-H_d^{-1}(\mu_d F_{Y_{11}|D=d})}{\lambda_d-1}\right)\right),\\ \overline{T}_d &= M_0\left(m_1\left(\frac{\lambda_d F_{Y_{01}|D=d}-H_d^{-1}(\mu_d F_{Y_{11}|D=d}+(1-\mu_d))}{\lambda_d-1}\right)\right).\end{aligned}$$

When $P(D_{00}=d)>0$, we can bound $F_{Y_{11}(d)|C}$ by $C_d(\underline{T}_d)$ and $C_d(\overline{T}_d)$. These bounds can however be improved by remarking that $F_{Y_{11}(d)|C}$ is increasing. Therefore, we define our bounds as:

$$\begin{aligned}\underline{B}_d(y) &= \sup_{y'\leq y} C_d\left(\underline{T}_d\right)(y'),\\ \overline{B}_d(y) &= \inf_{y'\geq y} C_d\left(\overline{T}_d\right)(y').\end{aligned}\qquad (4)$$

When $P(D_{00}=d)=0$, the bounds on $F_{Y_{11}(d)|C}$ are much simpler. We simply bound $F_{Y_{11}(1)|(2d-1)V\geq(2d-1)v_0(0)}$ by 0 and 1, which yields

$$\underline{B}_d(y)=M_0\left(\frac{P(D_{10}=d)-P(D_{11}=d)F_{Y_{11}|D=d}}{P(D_{10}=d)-P(D_{11}=d)}\right),\quad \overline{B}_d(y)=m_1\left(\frac{-P(D_{11}=d)F_{Y_{11}|D=d}}{P(D_{10}=d)-P(D_{11}=d)}\right).$$

For $d = 0$, the bounds are actually trivial since $\underline{B}_0(y) = 0$ and $\overline{B}_1(y) = 1$.

Another important case where the bounds take a simple form is when $P(D_{10} = 1) = 0$. In this case, one can check that

$$\underline{B}_1 = \overline{B}_1 = F_{Y_{11}|D=1}.$$

This is because in such situations, $F_{Y_{11}(1)|C} = F_{Y_{11}|D=1}$ as shown in Equation (3).

Theorem 3.3 proves that $\underline{B}_d$ and $\overline{B}_d$ are indeed bounds for $F_{Y_{11}(d)|C}$. We also consider the issue of whether these bounds are sharp or not. Hereafter, we say that $\underline{B}_d$ is sharp (and similarly for $\overline{B}_d$) if there exists a sequence of cdf $(G_k)_{k\in\mathbb{N}}$ such that supposing $F_{Y_{11}(d)|C} = G_k$ is compatible with both the data and the model, and for all y, $\lim_{k\to\infty} G_k(y) = \underline{B}_d(y)$. We establish that $\underline{B}_d$ and $\overline{B}_d$ are sharp under Assumption 7 below. Note that this assumption is testable from the data.

Assumption 7 *(Increasing bounds)*

For $(d, g, t) \in \{0, 1\}^3$, $F_{Y_{gt}|D=d}$ is continuously differentiable, with positive derivative on $\overset{\circ}{\mathcal{S}}(Y)$. Moreover, either (i) $P(D_{00} = d) = 0$ or (ii) $\underline{T}_d$, $G_d(\underline{T}_d)$ and $C_d(\underline{T}_d)$ (resp. $\overline{T}_d$, $G_d(\overline{T}_d)$ and $C_d(\overline{T}_d)$) are increasing.

Theorem 3.3 *If Assumptions 1-5 hold, we have*

$$\underline{B}_d(y) \leq F_{Y_{11}(d)|C}(y) \leq \overline{B}_d(y).$$

Moreover, if Assumption 7 holds, $\underline{B}_d(y)$ and $\overline{B}_d(y)$ are sharp.

The intuition underlying the sharpness result goes as follows. Let $\mathcal{T}_d$ be the set of all functions T_d increasing and included between 0 and 1 such that $G_d(T_d)$ and $C_d(T_d)$ are also increasing and included between 0 and 1. $\mathcal{T}_0$ is the set of all cdf $F_{Y_{01}(0)|TC}$ such that $F_{Y_{01}(0)|V<v_0(0)}$ and $F_{Y_{11}(0)|C}$ are cdf. This suggests that $\mathcal{T}_d$ is the set of all candidates for $F_{Y_{01}(d)|TC}$ that can be rationalized by the data and the model. Now, assume that there exists T_0^- and T_0^+ in $\mathcal{T}_0$ such that $T_0^- \leq T_0 \leq T_0^+$ for every $T_0 \in \mathcal{T}_0$. When $\lambda_0 > 1$, $G_0(.)$ is decreasing in T_0, which implies that $C_0(.)$ is also decreasing in T_0. Therefore, in such instances the sharp lower bound of $C_0(.)$ is equal to $C_0(T_0^+)$, while the sharp upper bound is equal to $C_0(T_0^-)$. Moreover, when $\lambda_0 > 1$, it appears after some algebra that T_0, $G_0(T_0)$ and $C_0(T_0)$ are all included between 0 and 1 if and only if $\overline{T}_0 \leq T_0 \leq \underline{T}_0$. Therefore, if $\underline{T}_0$, $G_0(\underline{T}_0)$ and $C_0(\underline{T}_0)$ are increasing, $\underline{T}_0$ is in $\mathcal{T}_0$ and $T_0^+ = \underline{T}_0$. This implies that $\underline{B}_0 = C_0(\underline{T}_0)$ is sharp under Assumption 7. When $\lambda_0 < 1$, a similar reasoning also shows that $\underline{B}_0 = C_0(\underline{T}_0)$ is sharp under Assumption 7.

Interestingly, when $[\underline{y}, \overline{y}] = (-\infty, +\infty)$ $\underline{B}_0$ and $\overline{B}_0$ are proper cdf when $\lambda_0 > 1$, but are defective when $\lambda_0 < 1$. More precisely, $\lim_{y\to+\infty} \underline{B}_0(y) < 1$ and $\lim_{y\to-\infty} \overline{B}_0(y) > 0$. The reason for this asymmetry is that when $\lambda_0 < 1$, time compliers belong to the group of treated observations in the control $\times$ period 1 cell (cf. Table 2). Therefore, their $Y(0)$ is not observed in period 1, and the data does not impose any restriction on $F_{Y_{01}(0)|TC}$: it could be equal to 0 or to 1, hence the defective bounds. On the contrary, when $\lambda_0 > 1$, time compliers belong to the group of untreated observations in the control $\times$ period 1 cell. Moreover, under Assumption 3, we know that they account for $100(1 - 1/\lambda_0)\%$ of this group. Consequently, the data imposes some restrictions on $F_{Y_{01}(0)|TC}$. For instance, we must have

$$F_{Y_{01}|D=0,Y\geq\alpha} \leq F_{Y_{01}(0)|TC} \leq F_{Y_{01}|D=0,Y\leq\beta},$$

where $\alpha = F^{-1}_{Y_{01}|D=0}(1/\lambda_0)$ and $\beta = F^{-1}_{Y_{01}|D=0}(1 - 1/\lambda_0)$. The cdf of time compliers cannot be below the one of the $100(1 - 1/\lambda_0)\%$ of observations with highest Y of this group, and cannot be above the one of the $100(1 - 1/\lambda_0)\%$ of observations with lowest Y of this group. $\underline{B}_0$ and $\overline{B}_0$ are trimming bounds in the spirit of Horowitz & Manski (1995) when $\lambda_0 > 1$, but not when $\lambda_0 < 1$, which is the reason why they are defective then.

Another interesting asymmetry is that $\underline{B}_1$ and $\overline{B}_1$ are always proper cdf, while we could have expected them to be defective when $\lambda_0 > 1$, because then time compliers are untreated in period 1, so their $Y(1)$ is unobserved. This second asymmetry stems from the fact that when $\lambda_0 > 1$, time compliers do not belong to our population of compliers ($C = IC \setminus TC$), while when $\lambda_0 < 1$, time compliers are included within our population of interest ($C = IC \cup TC$). Setting $F_{Y_{01}(1)|TC}(y) = 0$ does not imply that $\lim_{y\to+\infty} F_{Y_{11}(1)|C}(y) < 1$ when $TC \cap C$ is empty, while setting $F_{Y_{01}(0)|TC}(y) = 0$ yields $\lim_{y\to+\infty} F_{Y_{11}(0)|C}(y) < 1$ when $TC \subset C$.

Finally, one can check that $\lim_{y\to\overline{y}} \underline{B}_0(y) = M_0\left(\frac{P(D_{10}=0)H_0(\lambda_0)-P(D_{11}=0)}{P(D_{10}=0)-P(D_{11}=0)}\right)$, while $\lim_{y\to\underline{y}} \overline{B}_0(y) = m_1\left(\frac{P(D_{10}=0)H_0(1-\lambda_0)}{P(D_{10}=0)-P(D_{11}=0)}\right)$. Under Assumption 5, the first limit is strictly greater than 0 and the second one is strictly lower than 1, which implies that our two bounds are non trivial. If Assumption 5 is violated, at least one of our two bounds is trivial, which implies that for every quantile treatment effect one of our two bounds will either be $+\infty$ or $-\infty$.

A consequence of Theorem 3.3 is that QTE and LATE are partially identified when $P(D_{00} = 0) \neq P(D_{01} = 0)$ or $P(D_{00} = 0) \in \{0, 1\}$. The bounds are given in the following corollary. To ensure that the bounds on the LATE are well defined, we impose the following technical condition.

Assumption 8 *(Existence of moments)*

$\int |y|d\overline{B}_1(y) < +\infty$ and $\int |y|d\underline{B}_1(y) < +\infty$.[3]

[3] $\int |y|d\overline{B}_1(y)$ is the integral of the absolute value function with respect to the probability measure ν defined on $[\underline{y}, \overline{y}]$ and generated by $\overline{B}_1$. The same holds for $\int |y|d\underline{B}_1(y)$, $\int |y|d\overline{B}_0(y)$ and $\int |y|d\underline{B}_0(y)$. Because we may have $\lim_{y\to\underline{y}} \overline{B}_0(y) > 0$ or $\lim_{y\to\overline{y}} \overline{B}_0(y) < 1$, ν may admit a mass at $\underline{y}$ or $\overline{y}$.

Corollary 3.4 *If Assumptions 1-5 and 8 hold and $P(D_{00}=0) \neq P(D_{01}=0)$, Δ and τ_q are partially identified, with*

$$\int y d\overline{B}_1(y) - \int y d\underline{B}_0(y) \leq \Delta \leq \int y d\underline{B}_1(y) - \int y d\overline{B}_0(y),$$

$$\max(\overline{B}_1^{-1}(q), \underline{y}) - \min(\underline{B}_0^{-1}(q), \overline{y}) \leq \tau_q \leq \min(\underline{B}_1^{-1}(q), \overline{y}) - \max(\overline{B}_0^{-1}(q), \underline{y}).$$

Moreover, suppose that Assumption 7 holds. Then

- *If $\lambda_0 > 1$ or $E(|Y_{11}(0)|\,|C) < +\infty$, the bounds on Δ are sharp.*
- *If $\lambda_0 > 1$ or for $d \in \{0,1\}$, $\underline{B}_d(y) = q$ and $\overline{B}_d(y) = q$ admit a unique solution, the bounds on τ_q are sharp.*

When $\lambda_0 < 1$ and $\mathcal{S}(Y)$ is unbounded, the bounds on Δ are infinite, and some bounds on τ_q are also infinite. On the contrary, when $\lambda_0 > 1$ the bounds on τ_q are always finite, for every $q \in (0,1)$. The bounds on the LATE will also be finite in this case, as soon as $\underline{B}_0$ and $\overline{B}_0$ admit an expectation. Table 3 summarizes the situation.

Table 3: Finiteness of the bounds when $\underline{y} = -\infty$, $\overline{y} = +\infty$.

	$\lambda_0 < 1$	$\lambda_0 > 1$
$\underline{\tau}_q, q$ small	finite	finite
$\overline{\tau}_q, q$ small	$+\infty$	finite
$\underline{\tau}_q, q$ large	$-\infty$	finite
$\overline{\tau}_q, q$ large	finite	finite
$\underline{\Delta}$	$-\infty$	finite in general
$\overline{\Delta}$	$+\infty$	finite in general

q small means $0 < q < \underline{q}$ for a well chosen $\underline{q}$. Similarly,
q large means $\overline{q} < q < 1$ for a well chosen $\overline{q}$.

4 Extensions

4.1 Identification of a conditional IV-CIC model

We shall now consider a version of our IV-CIC model incorporating covariates, which we refer to as the conditional IV-CIC model. This will allow us to weaken our main identifying assumption and to outline two more strategies to recover point identification when $P(D_{00} = d) \neq P(D_{01} = d)$.

Let $\lambda_{d,x} = P(D_{01} = d|X = x)/P(D_{00} = d|X = x)$ and $\mu_{d,x} = P(D_{11} = d|X = x)/P(D_{10} = d|X = x)$. Assume that

$$Y(d) = h_d(U_d, T, X), \quad d \in \{0; 1\},$$

and substitute the following assumptions to Assumptions 1-5:

Assumption 9 *(Monotonicity 2)*

$h_d(u, t, x)$ is strictly increasing in u for all $(d, t, x) \in \{0, 1\}^2 \times \mathcal{S}(X)$.

Assumption 10 *(Latent index model for potential treatments 2)*

$D(z) = 1\{V \geq v_z(T, X)\}$ with $v_0(t, x) > v_1(t, x)$ for $(t, x) \in \{0; 1\} \times \mathcal{S}(X)$.

Assumption 11 *(Conditional time invariance)*

For $d \in \{0, 1\}$, $(U_d, V) \perp\!\!\!\perp T \,|\, G, X$.

Assumption 12 *(Data restrictions 2)*

1. *$\mathcal{S}(X_{gt}|D = d) = \mathcal{S}(X) = [\underline{x}, \overline{x}]$ with $(\underline{x}, \overline{x}) \in \overline{\mathbb{R}}^2$.*
2. *$\mathcal{S}(Y_{gt}|D = d, X = x) = \mathcal{S}(Y) = [\underline{y}, \overline{y}]$ with $(\underline{y}, \overline{y}) \in \overline{\mathbb{R}}^2$, for $(g, t, d, x) \in \{0; 1\}^3 \times \mathcal{S}(X)$.*
3. *$F_{Y_{gt}|D=d,X=x}$ is strictly increasing and continuous on $\mathcal{S}(Y)$, for $(g, t, d, x) \in \{0; 1\}^3 \times \mathcal{S}(X)$.*

Assumption 13 *(Changes in the treatment rates 2)*

For every $x \in \mathcal{S}(X)$,

1. *$P(D_{11} = 1|X = x) - P(D_{10} = 1|X = x) > 0$.*
2. *If $P(D_{00} = 0|X = x) > 0$, $F_{Y_{10}|D=0,X=x} \circ F^{-1}_{Y_{00}|D=0,X=x}(\lambda_{0,x}) > \mu_{0,x}$ and $F_{Y_{10}|D=0,X=x} \circ F^{-1}_{Y_{00}|D=0,X=x}(1 - \lambda_{0,x}) < 1 - \mu_{0,x}$.*

Incorporating covariates allows us to weaken the main identifying assumption of our model. When the distribution of some X evolves over time in the control or in the treatment group, Assumption 11 might be more credible than Assumption 3: if the distribution of X is not stable over time and X is correlated to (U_d, V), then the distribution of (U_d, V) might not be stable either.

The main results of the previous section extend to this conditional IV-CIC model. Firstly, notice that the distribution of X_{11} among compliers is identified under Assumptions 9-13, as shown in the next lemma.

Lemma 4.1 *Suppose that Assumptions 9-13 hold. Then,*

$$f_{X_{11}|C}(x) = \frac{\left[P(D_{11}=1|X=x) - P(D_{10}=1|X=x)\right] f_{X_{11}}(x)}{E\left[P(D_{11}=1|X) - P(D_{10}=1|X)|G=1,T=1\right]}$$

Then, the conditional IV-CIC assumptions imply that the IV-CIC assumptions are satisfied conditional on X, so one can prove conditional versions of Theorems 3.1 and 3.3. This implies that for every $d \in \{0,1\}$, $F_{Y_{11}(d)|C,X=x}(y)$ is point identified whenever $0 < P(D_{00} = d|X = x) = P(D_{01} = d|X = x)$, while it is partially identified otherwise. One can then integrate $F_{Y_{11}(d)|C,X=x}(y)$ or its bounds over the distribution of X_{11} among compliers to point or partially identify $F_{Y_{11}(d)|C}(y)$. This idea is formalized in the following theorem, in the point identified case. Hereafter, we let $Q_{d,x}(y) = F^{-1}_{Y_{01}|D=d,X=x} \circ F_{Y_{00}|D=d,X=x}(y)$.

Theorem 4.1 *Suppose Assumptions 9-13 hold. If $0 < P(D_{00} = d|X) = P(D_{01} = d|X)$ almost surely, the conditional distribution of potential outcomes on compliers is identified by*

$$F_{Y_{11}(d)|X=x,C}(y) = \frac{P(D_{10}=d|X=x)F_{Q_{d,x}(Y_{10})|D=d,X=x}(y) - P(D_{11}=d|X=x)F_{Y_{11}|D=d,X=x}(y)}{P(D_{10}=d|X=x) - P(D_{11}=d|X=x)}.$$

The overall distribution of potential outcomes among compliers is also identified.

This theorem is useful when $P(D_{00} = d) \neq P(D_{01} = d)$ but $P(D_{00} = d|X) = P(D_{01} = d|X) > 0$ almost surely, meaning that in the control group, the evolution of the treatment rate is entirely driven by a change in the distribution of X over time. Otherwise, we can of course obtain bounds, using a similar argument as in Theorem 3.3. The bounds are likely to be tighter than the unconditional ones if X drives most of the evolution of the treatment rate in the control group.

We also consider another assumption under which treatment effects are still point identified, even if the treatment rate evolves in some X cells of the control group.

Assumption 14 *(Strong conditional time invariance)*

$U_d \perp\!\!\!\perp T|G = 0, D(0) = d, X$

When $P(D_{00} = 0|X = x) = P(D_{01} = 0|X = x)$, Assumption 11 implies that Assumption 14 holds in the $X = x$ cell. But this is no longer true when $P(D_{00} = 0|X = x) \neq P(D_{01} = 0|X = x)$. Then, Assumption 14 requires that even though selection into treatment might evolve over time, treated (resp. untreated) observations in the control group have the same distribution of U_1 (resp. U_0) in period 0 and 1. This might be credible when the change in the treatment rate between the two periods is small, and X captures most of selection into treatment. If $P(D_{00} = 0|X = x) < P(D_{01} = 0|X = x)$, a sufficient condition for that to hold in the $X = x$ cell is

$$U_d \perp\!\!\!\perp V|G = 0, T = d, D(0) = d, X = x, \quad d \in \{0, 1\}.$$

This is reminiscent of the ignorability condition in Rosenbaum & Rubin (1983), even though we believe it is weaker. Ignorability states that selection is exogenous after controlling for X. Here we posit that the propensity for the treatment is exogenous after controlling for both X and $D(0)$.

Theorem 4.2 *Suppose Assumptions 9-13 and 14 hold. The conditional distributions of potential outcomes on compliers are identified by*

$$F_{Y_{11}(d)|X=x,C}(y) = \frac{P(D_{10} = d|X = x)F_{Q_{d,x}(Y)_{10}|D=d,X=x}(y) - P(D_{11} = d|X = x)F_{Y_{11}|D=d,X=x}(y)}{P(D_{10} = d|X = x) - P(D_{11} = d|X = x)}$$

for $d \in \{0, 1\}$. *The overall distribution of potential outcomes among compliers is also identified.*

4.2 Several periods and groups

Results of Section 3 can also be extended to settings with many groups. This will increase the chances that we can recover point identification, provide us with a test of our model, and at the very least tighten our bounds relative to the two groups case. If the treatment rate is stable in at least one group, which is likely to be the case with many groups, one can use it as a control group and point identify treatment effects in period 1 among compliers in every group in which the treatment rate changes between the two periods.[4] If there are several groups in which the treatment rate remains stable between the two periods, one can use either of those groups as a control for other groups. This provides us with a test of our IV-CIC model, as the quantile-quantile transforms of the outcome should be the same in all these control groups. Formally, $F^{-1}_{Y_{g0}|D=d} \circ F_{Y_{g1}|D=d}$ should not depend on g, for any g such that $P(D_{g1} = 0) = P(D_{g0} = 0)$. If the treatment rate changes in every group, then for each group we can derive bounds for the cdf of $Y(0)$ and $Y(1)$ among compliers using any other group satisfying Assumption 5 as a control group, and we can tighten the various bounds obtained by using intersection bounds (see Chernozhukov et al., 2013).

The previous results can also be extended to settings with many time periods, which will increase even more more the chances of recovering point identification. With more than two

[4]For groups in which the treatment rate diminishes over the two periods, one can just switch labels and consider $1 - D$ as the treatment variable. As explained before, the first condition in Assumption 5 is just a normalization. Doing this, we recover the opposite of the LATE and QTE, for a population of compliers defined differently (the individuals who switch from $D = 1$ to $D = 0$ either because of time or because of the instrument).

periods, if the treatment rate is stable in at least one group between period $t-k$ and t for some k, one can use it as a control group and point identify treatment effects in period t among compliers in every group in which the treatment rate changes between $t-k$ and t. However, our time invariance assumption might be less credible when the number of time periods grows, since groups are more likely to evolve over a long time period.

4.3 Testability

We show now that our IV-CIC model is testable. We focus here on the unconditional model described in Section 2, but similar implications could be obtained for the conditional models considered above. For every $y \leq y'$ in $\mathcal{S}(Y)^2$, let

$$I_d(y,y') = [\min(\underline{T}_d(y), \overline{T}_d(y)), \max(\underline{T}_d(y'), \overline{T}_d(y'))],$$

with the convention that $I_d(y,y') = \emptyset$ if

$$\min(\underline{T}_d(y), \overline{T}_d(y)) > \max(\underline{T}_d(y'), \overline{T}_d(y')).$$

Theorem 4.3 *If Assumption 4 holds, we reject Assumptions 1-3 together if for some $d \in \{0;1\}$, one of the two following statements holds:*

1. *For some $y_0 \leq y_1$ in $\mathcal{S}(Y)^2$, $I_d(y_0,y_1) = \emptyset$.*
2. *For some $y_0 < y_1$ in $\mathcal{S}(Y)^2$, $I_d(y_0,y_1) \neq \emptyset$ but for every $t_0 \leq t_1$ in $I_d(y_0,y_1)^2$,*

$$\begin{aligned} & \frac{P(D_{10}=d)H_d \circ (\lambda_d F_{Y_{01}|D=d}(y_1) + (1-\lambda_d)t_1) - P(D_{11}=d)F_{Y_{11}|D=d}(y_1)}{P(D_{10}=d) - P(D_{11}=d)} \\ < \quad & \frac{P(D_{10}=d)H_d \circ (\lambda_d F_{Y_{01}|D=d}(y_0) + (1-\lambda_d)t_0) - P(D_{11}=d)F_{Y_{11}|D=d}(y_0)}{P(D_{10}=d) - P(D_{11}=d)}. \end{aligned} \quad (5)$$

Theorem 4.3 provides a theoretical test of the model. When the treatment rate does not change in the control group, i.e. when $\lambda_d = 1$, the two testable implications of the IV-CIC model reduce to having that

$$\frac{P(D_{10}=d)H_d \circ F_{Y_{01}|D=d}(y) - P(D_{11}=d)F_{Y_{11}|D=d}(y)}{P(D_{10}=d) - P(D_{11}=d)}$$

is increasing. Therefore, in such instances, our test is similar to the one developed by Kitagawa (2013) for the instrumental variable model in Angrist et al. (1996) with binary treatment and instrument. In their model, the cdf of compliers can also be written as the difference between two increasing functions, and thus may not be increasing (see Imbens & Rubin, 1997).

On the contrary, when the treatment rate changes in the control group, our test of the IV-CIC model is slightly different. In such instances, we can reject the model when $\mathcal{T}_d$ is empty for some $d \in \{0,1\}$, i.e. when there is no function T_d such that $G_d(T_d)$ and $C_d(T_d)$ are also cdf, while such a function should exist under Assumptions 1-3 as shown in Lemma 3.1. We give two sufficient conditions under which $\mathcal{T}_d$ is empty. To understand them, remark that $I_d(y,y')^2$ includes the set of all possible values for $T_d(y)$ and $T_d(y')$ such that $T_d(y)$, $T_d(y')$, $G_d(T_d)(y)$, $C_d(T_d)(y)$, $G_d(T_d)(y')$, and $C_d(T_d)(y')$ are included between 0 and 1. If $I_d(y_0,y_1)$ is empty for some $y_0 \leq y_1$, $\mathcal{T}_d$ must be empty. If point 2 holds, $\mathcal{T}_d$ is also empty because it is not possible to define $T_d(y_0)$ and $T_d(y_1)$ such that $0 \leq T_d(y_0) \leq T_d(y_1) \leq 1$, $G_d(T_d)(y_0)$, $G_d(T_d)(y_1)$, $C_d(T_d)(y_0)$ and $C_d(T_d)(y_1)$ are included between 0 and 1 and $C_d(T_d)(y_0) \leq C_d(T_d)(y_1)$.

The test presented in point 2 is much simpler to implement when $\lambda_0 < 1$ than when $\lambda_0 > 1$. When $\lambda_0 < 1$,

$$\frac{P(D_{10}=d)H_d \circ (\lambda_d F_{Y_{01}|D=d}(y) + (1-\lambda_d)t) - P(D_{11}=d)F_{Y_{11}|D=d}(y)}{P(D_{10}=d) - P(D_{11}=d)}$$

is increasing in t. Therefore, one necessary and sufficient condition for inequality (5) to hold in this case is that there exists $y_0 \leq y_1$ in $\mathcal{S}(Y)^2$ such that for some $d \in \{0,1\}$,

$$\begin{aligned}&\frac{P(D_{10}=d)H_d \circ (\lambda_d F_{Y_{01}|D=d}(y_1) + (1-\lambda_d)\max(\underline{T}_d(y_1), \overline{T}_d(y_1))) - P(D_{11}=d)F_{Y_{11}|D=d}(y_1)}{P(D_{10}=d) - P(D_{11}=d)}\\ < \quad &\frac{P(D_{10}=d)H_d \circ (\lambda_d F_{Y_{01}|D=d}(y_0) + (1-\lambda_d)\min(\underline{T}_d(y_0), \overline{T}_d(y_0))) - P(D_{11}=d)F_{Y_{11}|D=d}(y_0)}{P(D_{10}=d) - P(D_{11}=d)}.\end{aligned}$$

In contrast, when $\lambda_0 > 1$,

$$\frac{P(D_{10}=d)H_d \circ (\lambda_d F_{Y_{01}|D=d}(y) + (1-\lambda_d)t) - P(D_{11}=d)F_{Y_{11}|D=d}(y)}{P(D_{10}=d) - P(D_{11}=d)}$$

is decreasing in t. Then, one necessary and sufficient condition for inequality (5) to hold in this case is that there exists $y_0 < y_1$ in $\mathcal{S}(Y)^2$ such that for some $d \in \{0,1\}$ and for every t in $I(y_0,y_1)$,

$$\begin{aligned}&\frac{P(D_{10}=d)H_d \circ (\lambda_d F_{Y_{01}|D=d}(y_1) + (1-\lambda_d)t) - P(D_{11}=d)F_{Y_{11}|D=d}(y_1)}{P(D_{10}=d) - P(D_{11}=d)}\\ < \quad &\frac{P(D_{10}=d)H_d \circ (\lambda_d F_{Y_{01}|D=d}(y_0) + (1-\lambda_d)t) - P(D_{11}=d)F_{Y_{11}|D=d}(y_0)}{P(D_{10}=d) - P(D_{11}=d)}.\end{aligned}$$

When $\lambda_0 < 1$, for every $y_0 < y_1$ the test will amount to assessing inequality (5) for only one value of (t_0,t_1). When $\lambda_0 > 1$, for every $y_0 < y_1$ the test will amount to assessing Inequality (5) for an infinity of (t_0,t_1), unless $I(y_0,y_1)$ reduces to a point.

5 Inference

In this section, we develop inference on LATE and QTE in the point and partially identified cases. In both cases, we impose the following conditions.

Assumption 15 *(Independent and identically distributed observations)*

$(Y_i, D_i, G_i, T_i)_{i=1,...,n}$ *are i.i.d.*

Assumption 16 *(Technical conditions for inference 1)*

$\mathcal{S}(Y)$ *is a bounded interval* $[\underline{y}, \overline{y}]$. *Moreover, for all* $(d, g, t) \in \{0,1\}^3$, $F_{dgt} = F_{Y_{gt}|D=d}$ *and* $F_{Y_{11}(d)|C}$ *are continuously differentiable with strictly positive derivatives on* $[\underline{y}, \overline{y}]$.

Athey & Imbens (2006) impose a condition similar to Assumption 16 when studying the asymptotic properties of their estimator.

We first consider the point identified case, which corresponds either to $0 < P(D_{00} = 0) = P(D_{01} = 0) < 1$ under Assumptions 1-5, or to $P(D_{00} = 0) = P(D_{01} = 0) \in \{0,1\}$ under Assumptions 1-6. For simplicity, we focus hereafter on the first case but the estimator and its asymptotic properties are completely similar in the second case. Let $\widehat{F}_{dgt}$ (resp. $\widehat{F}^{-1}_{dgt}$) denote the empirical cdf (resp. quantile function) of Y on the subsample $\{i : D_i = d, G_i = g, T_i = t\}$ and $\widehat{Q}_d = \widehat{F}^{-1}_{d01} \circ \widehat{F}_{d00}$. We also let $\mathcal{I}_{gt} = \{i : G_i = g, T_i = t\}$ and n_{gt} denote the size of $\mathcal{I}_{gt}$ for all $(d, g) \in \{0,1\}^2$. Our estimator of the LATE is

$$\widehat{\Delta} = \frac{\frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{11}} Y_i - \frac{1}{n_{10}}\sum_{i\in\mathcal{I}_{10}} \widehat{Q}_{D_i}(Y_i)}{\frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{11}} D_i - \frac{1}{n_{10}}\sum_{i\in\mathcal{I}_{10}} D_i}$$

Let $\widehat{P}(D_{gt} = d)$ be the proportion of subjects with $D = d$ in the sample $\mathcal{I}_{gt}$, let $\widehat{H}_d = \widehat{F}_{d10} \circ \widehat{F}^{-1}_{d00}$, and let

$$\widehat{F}_{Y_{11}(d)|C} = \frac{\widehat{P}(D_{01} = d)\widehat{H}_d \circ \widehat{F}_{d01} - \widehat{P}(D_{11} = d)\widehat{F}_{d11}}{\widehat{P}(D_{10} = d) - \widehat{P}(D_{11} = d)}.$$

Our estimator of the QTE of order q for compliers is

$$\widehat{\tau}_q = \widehat{F}^{-1}_{Y_{11}(1)|C}(q) - \widehat{F}^{-1}_{Y_{11}(0)|C}(q).$$

We say hereafter that an estimator $\widehat{\theta}$ of a parameter θ is root-n consistent and asymptotically normal if there exists Σ such that $\sqrt{n}(\widehat{\theta}-\theta) \xrightarrow{L} \mathcal{N}(0, \Sigma)$. Theorem 5.1 below shows that both $\widehat{\Delta}$ and $\widehat{\tau}_q$ are root-n consistent and asymptotically normal. Because the asymptotic variances take complicated expressions, we consider the bootstrap for inference. For any statistic T, we let T^* denote its bootstrap counterpart. For any root-n consistent statistic $\widehat{\theta}$ estimating consistently the parameter θ, we say that the bootstrap is consistent if with probability one and conditional on the sample, $\sqrt{n}(\widehat{\theta}^* - \widehat{\theta})$ converges to the same distribution as the limit distribution of $\sqrt{n}(\widehat{\theta} - \theta)$.[5] Theorem 5.1 also shows that bootstrap confidence intervals are asymptotically valid.

[5]See, e.g., van der Vaart (2000), Section 23.2.1, for a formal definition of conditional convergence.

Theorem 5.1 *Suppose that Assumptions 1-5, 15-16 hold and* $0 < P(D_{00} = 0) = P(D_{01} = 1) < 1$*. Then* $\widehat{\Delta}$ *and* $\widehat{\tau}_q$ *are root-n consistent and asymptotically normal. Moreover, the bootstrap is consistent for both* $\widehat{\Delta}$ *and* $\widehat{\tau}_q$*.*

We now turn to the partially identified case. First, suppose that $0 < \widehat{P}(D_{00} = 0) < 1$ and $0 < \widehat{P}(D_{10} = 0) < 1$. Let $\widehat{\lambda}_d = \frac{\widehat{P}(D_{01}=d)}{\widehat{P}(D_{00}=d)}$, $\widehat{\mu}_d = \frac{\widehat{P}(D_{11}=d)}{\widehat{P}(D_{10}=d)}$ and define

$$\begin{aligned}
\widehat{\underline{T}}_d &= M_0\left(m_1\left(\frac{\widehat{\lambda}_d\widehat{F}_{Y_{01}|D=d} - \widehat{H}_d^{-1}(\widehat{\mu}_d\widehat{F}_{Y_{11}|D=d})}{\widehat{\lambda}_d - 1}\right)\right),\\
\widehat{\overline{T}}_d &= M_0\left(m_1\left(\frac{\widehat{\lambda}_d\widehat{F}_{Y_{01}|D=d} - \widehat{H}_d^{-1}(\widehat{\mu}_d\widehat{F}_{Y_{11}|D=d} + (1-\widehat{\mu}_d))}{\widehat{\lambda}_d - 1}\right)\right),\\
\widehat{G}_d(T) &= \widehat{\lambda}_d\widehat{F}_{Y_{01}|D=d} + (1-\widehat{\lambda}_d)T,\\
\widehat{C}_d(T) &= \frac{\widehat{P}(D_{10}=d)\widehat{H}_d\circ\widehat{G}_d(T) - \widehat{P}(D_{11}=d)\widehat{F}_{Y_{11}|D=d}}{\widehat{P}(D_{10}=d) - \widehat{P}(D_{11}=d)}.
\end{aligned}$$

To estimate bounds for $F_{Y_{11}(d)|C}$, we use

$$\widehat{\underline{B}}_d(y) = \sup_{y'\le y}\widehat{C}_d\left(\widehat{\underline{T}}_d\right)(y'),\quad \widehat{\overline{B}}_d(y) = \inf_{y'\ge y}\widehat{C}_d\left(\widehat{\overline{T}}_d\right)(y').$$

Therefore, to estimate bounds for the LATE and QTE, we use

$$\widehat{\underline{\Delta}} = \int yd\widehat{\overline{B}}_1(y) - \int yd\widehat{\underline{B}}_0(y),\quad \widehat{\overline{\Delta}} = \int yd\widehat{\underline{B}}_1(y) - \int yd\widehat{\overline{B}}_0(y),$$

$$\widehat{\underline{\tau}}_q = \widehat{\overline{B}}_1^{-1}(q) - \widehat{\underline{B}}_0^{-1}(q),\quad \widehat{\overline{\tau}}_q = \widehat{\underline{B}}_1^{-1}(q) - \widehat{\overline{B}}_0^{-1}(q).$$

When $\widehat{P}(D_{00} = 0) \in \{0, 1\}$ or $\widehat{P}(D_{10} = 0) \in \{0, 1\}$, the bounds on Δ and τ_q are defined similarly, but instead of $\widehat{\underline{B}}_d$ and $\widehat{\overline{B}}_d$, we use the empirical counterparts of the bounds on $F_{Y_{11}(d)|C}$ given by Equation (4).

Let $B_\Delta = (\underline{\Delta}, \overline{\Delta})$ and $B_{\tau_q} = (\underline{\tau}_q, \overline{\tau}_q)'$, and let $\widehat{B}_\Delta$ and $\widehat{B}_{\tau_q}$ be the corresponding estimators. Theorem 5.2 below establishes the asymptotic normality and the validity of the bootstrap for both $\widehat{B}_\Delta$ and $\widehat{B}_{\tau_q}$, for $q \in \mathcal{Q} \subset (0, 1)$, where $\mathcal{Q}$ is defined as follows. First, when $P(D_{00} = 0) \in \{0, 1\}$, $P(D_{10} = 0) = 1$,[6] or $\lambda_0 > 1$, we merely let $\mathcal{Q} = (0, 1)$. When $\lambda_0 < 1$, we have to exclude small and large q from $\mathcal{Q}$. This is because the (true) bounds put mass at the boundaries $\underline{y}$ or $\overline{y}$ of the support of Y. Similarly, the estimated bounds put mass on the estimated boundaries, which must be estimated. Because estimated boundaries typically have non-normal limit distribution, the asymptotic distribution of the bounds of the estimated QTE will also be non-normal. We thus restrict ourselves to $(\underline{q}, \overline{q})$, with $\underline{q} = \overline{B}_0(\underline{y})$ and $\overline{q} = \underline{B}_0(\overline{y})$.

[6]Assumption 4 rules out $P(D_{10} = 0) = 0$

Another issue is that the bounds might be irregular at some $q \in (0,1)$, because they include in their definitions the kinked functions M_0 and m_1.[7] Let

$$q_1 = \frac{\mu_1 F_{Y_{11}|D=1} \circ F^{-1}_{Y_{01}|D=1}(\frac{1}{\lambda_1}) - 1}{\mu_1 - 1}, \quad q_2 = \frac{\mu_1 F_{Y_{11}|D=1} \circ F^{-1}_{Y_{01}|D=1}(1 - \frac{1}{\lambda_1})}{\mu_1 - 1}$$

denote the two points at which the bounds can be kinked. When $\lambda_0 < 1$, we restrict ourselves to $\mathcal{Q} = (\underline{q}, \overline{q}) \backslash \{q_1, q_2\}$. Note that q_1 and q_2 may not belong to $(\underline{q}, \overline{q})$, depending on λ_1 and μ_1, so that $\mathcal{Q}$ may in fact be equal to $(\underline{q}, \overline{q})$.

Theorem 5.2 relies on the following technical assumption, which involves the bounds rather than the true cdf since we are interested in estimating these bounds. Note that the strict monotonicity requirement is only a slight reinforcement of Assumption 7.

Assumption 17 *(Technical conditions for inference 2)*

For $d \in \{0,1\}$, the sets $\underline{\mathcal{S}}_d = [\underline{B}_d^{-1}(\underline{q}), \underline{B}_d^{-1}(\overline{q})] \cap \mathcal{S}(Y)$ and $\overline{\mathcal{S}}_d = [\overline{B}_d^{-1}(\underline{q}), \overline{B}_d^{-1}(\overline{q})] \cap \mathcal{S}(Y)$ are not empty. The bounds $\underline{B}_d$ and $\overline{B}_d$ are strictly increasing on $\underline{\mathcal{S}}_d$ and $\overline{\mathcal{S}}_d$. Their derivative, whenever they exist, are strictly positive.

Theorem 5.2 *Suppose that Assumptions 1-5, 7, 15-17 hold and $q \in \mathcal{Q}$. Then $\widehat{B}_\Delta$ and $\widehat{B}_{\tau_q}$ are root-n consistent and asymptotically normal. Moreover, the bootstrap is consistent for both.*

To construct confidence intervals of level $1-\alpha$ for Δ (resp. τ_q), one can use the lower bound of the two-sided (bootstrap) confidence interval of level $1-\alpha$ of $\underline{\Delta}$ (resp. $\underline{\tau_q}$), and the upper bound of the two-sided (bootstrap) confidence interval of $\overline{\Delta}$ (resp. $\overline{\tau_q}$). Such confidence intervals are asymptotically valid but conservative. Because $\underline{\Delta} < \overline{\Delta}$ (resp. $\underline{\tau_q} < \overline{\tau_q}$), a confidence interval on Δ (resp. τ_q) could alternatively be based on one-sided confidence intervals of level $1-\alpha$ on $\underline{\Delta}$ and $\overline{\Delta}$ (resp $\underline{\tau_q}$ and $\overline{\tau_q}$).[8]

Those results can easily be generalized to the conditional IV-CIC estimands presented in Section 4.1, provided covariates are discrete. Let Δ_x denote the LATE among compliers in the $X = x$ cell. Let $\widehat{\Delta}_x$ denote its plug-in estimator in cells such that $0 < P(D_{01} = 1|X = x) = P(D_{00} = 1|X = x) < 1$, and let $\widehat{B}_{\Delta_x} = \left(\widehat{\underline{\Delta}}_x, \widehat{\overline{\Delta}}_x\right)'$ denote the plug-in estimators of its bounds when $P(D_{01} = 1|X = x) \neq P(D_{00} = 1|X = x)$. One can generalize Theorems

[7]This problem does not arise when $\lambda_0 > 1$. Kinks are possible only at 0 or 1 in this case.

[8]As shown in Imbens & Manski (2004), such confidence intervals suffer however from a lack of uniformity, since their coverage rate falls below the nominal level when one gets close to point identification, i.e. when $\lambda_d \to 1$. The solutions to this problem suggested by Imbens & Manski (2004) or Stoye (2009) require that bounds converge uniformly towards normal distributions. In Theorem 5.2, we only show pointwise convergence, not uniform convergence. Uniform convergence is likely to fail for QTE because of the possible kinks of $\underline{B}_d$ and $\overline{B}_d$ at the points q_1 or q_2, which themselves depend on the underlying data generating process.

5.1 and 5.2 to show that $\widehat{\Delta}_x$ and $\widehat{B}_{\Delta_x}$ are root-n consistent, asymptotically normal, and the bootstrap is consistent for both of them. If $0 < P(D_{01} = 1|X) = P(D_{00} = 1|X) < 1$ almost surely, we define $\widehat{\Delta}$ as a weighted average of $\widehat{\Delta}_x$, using the sample equivalent of the density defined in Lemma 4.1 as weights. Those estimated weights are consistent and asymptotically normal. One can therefore show that $\widehat{\Delta}$ is root-n consistent, asymptotically normal, and the bootstrap is consistent for it. If for some x we do not have $0 < P(D_{01} = 1|X = x) = P(D_{00} = 1|X = x) < 1$, we define $\underline{\widehat{\Delta}}$ as a weighted average of $\widehat{\Delta}_x$ or $\underline{\widehat{\Delta}}_x$ (depending on whether $0 < P(D_{01} = 1|X = x) = P(D_{00} = 1|X = x) < 1$ or not in each $X = x$ cell), using the same weights as above. Similarly, we define $\overline{\widehat{\Delta}}$ as a weighted average of $\widehat{\Delta}_x$ or $\overline{\widehat{\Delta}}_x$. Let $\widehat{B}_\Delta = \left(\underline{\widehat{\Delta}}, \overline{\widehat{\Delta}}\right)'$. Since the estimated weights are consistent, one can also show that $\widehat{B}_\Delta$ is root-n consistent, asymptotically normal, and the bootstrap is consistent for it. A similar reasoning shows that estimates of QTE derived from the conditional IV-CIC model are root-n consistent, asymptotically normal, and the bootstrap is consistent for them.

So far, we have implicitly considered that we know whether point identification or partial identification holds, which is not the case in practice. This is an important issue, since the estimators and the way confidence intervals are constructed differ in the two cases. Abstracting from extreme cases where $P(D_{gt} = d) = 0$, testing point identification is simply equivalent to testing $\lambda_0 = 1$ versus $\lambda_0 \neq 1$. $\widehat{\lambda}_0$ is a root-n consistent estimator. Therefore, one can conduct asymptotically valid inference by checking first whether $|\widehat{\lambda}_0 - 1| \leq c_n$, with $(c_n)_{n\in\mathbb{N}}$ a sequence satisfying $c_n \to 0$, $\sqrt{n}c_n \to \infty$, and then applying either the point or partially identified framework. Such a pretest ensures that asymptotically, the probability of conducting inference under the wrong maintained assumption vanishes to 0. Inference following this pretest is therefore valid. Such a prestet is similar to procedures recently developed for inequality selection in moment inequality models (see for instance Andrews & Soares, 2010). In the conditional IV-CIC model, one must test whether $P(D_{01} = 1|X) = P(D_{00} = 1|X)$ almost surely in order to assess whether one can use Theorem 4.1. If X is discrete, one can test for this by running a saturated regression of D on X and T among control group observations, and testing for the joint significance of the $T \times X$ coefficients. The resulting F-statistic should also be compared with a sequence $(c_n)_{n\in\mathbb{N}}$ satisfying $c_n \to 0$ and $\sqrt{n}c_n \to \infty$ and not to its standard critical values, to account for pretesting. In the moment inequality literature, the choice of $c_n = \ln(\ln(n))/\sqrt{n}$ has often been advocated (see Andrews & Soares, 2010), so we will stick to it in our application.

6 Applications

6.1 Property rights and labor supply

Between 1996 and 2003, the Peruvian government issued property titles to over 1.2 million urban households, the largest titling program targeted to urban squatters in the developing world. Field (2007) examines the labor market effects of increases in tenure security resulting from the program. Tenure insecurity in Peru encompasses both fear of eviction by the government and fear of property theft by other residents. Such concerns might remove individuals from the labor force. In a nationwide survey of Peruvian households cited by the author, 47% of untitled households report keeping someone at home for property protection.

To isolate the causal effect of property rights, the author uses a survey conducted in 2000, and exploits two sources of variation in exposure to the titling program at that time. Firstly, this program took place at different dates in different neighborhoods. In 2000, it had approximately reached 50% of targeted neighborhoods. Secondly, it only impacted squatters, i.e. households without a property title prior to the program. The author can therefore construct four groups of households: squatters in neighborhoods reached by the program before 2000, squatters in neighborhoods reached by the program after 2000, non- squatters in neighborhoods reached by the program before 2000, and non- squatters in neighborhoods reached by the program after 2000. The share of households with a property title in each group is shown in Table 4.

Table 4: Share of households with a property right

	Reached after 2000	Reached before 2000
Squatters	0%	71%
Non-squatters	100%	100%

In Table 5 of the paper, the author estimates IV-DID regressions to capture the effect of having a property right on the total number of hours worked per week by the household. Whether the neighborhood was reached before or after 2000 plays the role of time, while squatters and non-squatters are the two groups. In what follows, we use the same data to measure the effect of property rights using our IV-CIC model instead of linear IV-DID regressions. As no squatters have a property right in neighborhoods reached after 2000, $P(D_{10} = 1) = 0$, so $F_{Y_{11}(1)|C}(y)$ is identified by $F_{Y_{11}|D=1|C}(y)$ as shown in Equation (3). Moreover, as 100% of non- squatters have a property title in the two groups of neighborhoods, $P(D_{00} = 1) = P(D_{10} = 1) = 1$. We can therefore use Theorem 3.2 to identify $F_{Y_{11}(0)|C}(y)$. The resulting estimates are displayed in Figure 1. $\widehat{F}_{Y_{11}(0)|C}$ stochastically dominates $\widehat{F}_{Y_{11}(1)|C}$, meaning that property rights have a positive impact on the number of hours worked, over the entire distribution of hours.

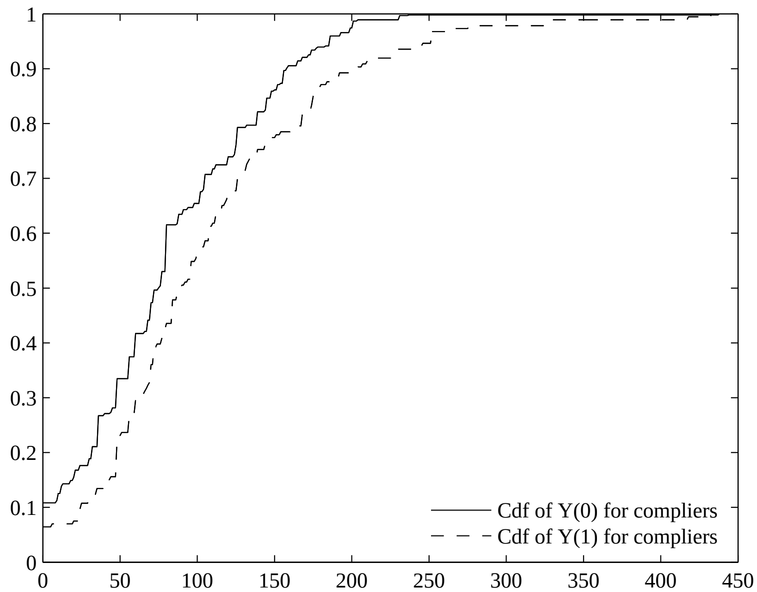

Figure 1: Estimated cdf of $Y(0)$ and $Y(1)$ for compliers.

As per our IV-CIC model, the LATE of titling on hours of work is equal to 23.3. This point estimate is 17% lower than the one we would have obtained through an IV-DID regression (27.2), and the difference between the two is statistically significant (p-value=0.02).[9] Interestingly, the quantile treatment effects on the level of the outcome are fairly constant, most of them being close to +20 hours of work per week. This implies that the effect of the treatment is highly heterogeneous in relative terms. Figure 2 shows quantile treatment effects on the logarithm of the outcome. Being granted a property title increases labor supply by more than 40% for households at the 25th percentile of the distribution of hours worked per week, and by 10% only for households at the 75th percentile. The difference between the two is marginally significant (p-value=0.12). An explanation for this pattern could go as follows. The main source of variation in hours worked per week at the household level is presumably the size of the household. In every household, only one household member has to stay home to look after the household's residence, irrespective of the household size. Being granted a property title therefore allows this household member to increase her labor supply, but has no effect on the

[9] Our IV-DID LATE does not match exactly the "Tilted" coefficient in the second column of Table 5 in Field (2007). We estimated the same regression as the author, but without control variables. This is to ensure that the resulting coefficient is comparable with our IV-CIC LATE, which is estimated without controls. Our IV-CIC model allows for discrete controls, but here the sample size is too small to include as many as the author did.

labor supply of other members. Knowing this pattern of heterogeneity might have substantial consequences on social choice. A utilitarian social planner will indeed be more prone to implementing a titling program with heterogeneous than with constant relative effects, provided utility of agents is concave in individual consumption.

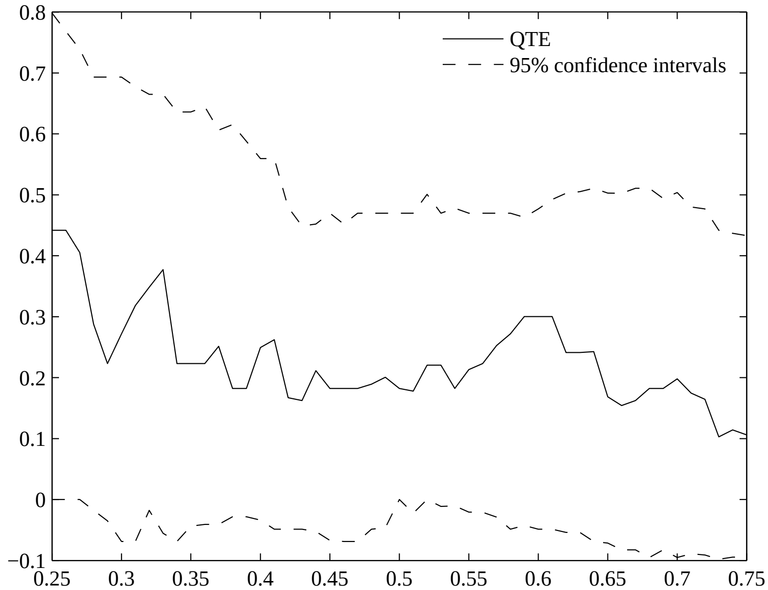

Figure 2: Estimated QTE on the logarithm of number of hours worked.

6.2 Returns to education and the Sekolah Dasar INPRES school construction program

In 1973, the Indonesian government launched a major school construction program, the so-called Sekolah Dasar INPRES program. It led to the construction of more than 61,000 primary schools between 1973-1974 and 1978-1979, an average of 2 schools per 1,000 children. Duflo (2001) uses the 1995 SUPAS census to measure the effect of this program on completed years of education in a first step, and returns to education in a second step. In what follows, we only consider the latter set of results.

There was substantial variation in treatment intensity across regions, as the government tried to allocate more schools to districts with low initial enrolment. The author thus constructs two groups of high and low program regions, by regressing the number of schools constructed on the number of children in each region. High treatment regions are those with a positive residual in that regression, as they received more schools than what their population predicts. Exposure to treatment also varied according to birth cohort: children between 2 and 6 in 1974 were exposed to the treatment as they were to enter primary school after the program was launched, while children between 12 and 17 in 1974 were not exposed as they were to have finished primary school by that time.

Number of years of education is larger for the second cohort in the two groups of regions, as schools were constructed in both groups. But the difference is larger in high treatment regions because more schools were constructed there. The author exploits this pattern to measure returns to education. She uses first a simple IV-DID regression in which birth cohort plays the role of the time variable, while low and high treatment regions are the two groups. The resulting coefficient, which we can infer from Table 3, is imprecisely estimated, so the author turns to richer specifications. All of them include cohort and region of birth fixed effects, so one can show that the resulting coefficient is a weighted average of Wald-DID across all possible pairs of regions and birth cohorts.

In what follows, we use the same data to measure returns to education using our IV-CIC model. As it does not allow for a multivariate treatment, we consider a dummy for whether an individual completed primary education as our treatment variable. The variable used in Duflo (2001) to construct completed years of education is a categorical variable (completed primary school, middle school...), with 9 categories. As the program was a primary school construction program, the larger increase in completed years of education in high program regions mostly comes from a larger increase in the share of individuals completing primary school. For instance, the share of individuals completing middle school did not evolve differently in the two groups of regions. Therefore, it seems that our binary treatment captures most of the variation in educational attainment induced by the program.

Table 5: Share of individuals completing primary school

	Older cohort	Younger cohort
High treatment regions	81.2%	90.0%
Low treatment regions	89.8%	94.3%

Table 5 shows that the share of individuals completing primary school increased more in high than in low treatment regions. As $0 < \widehat{P}(D_{10} = d) \neq \widehat{P}(D_{00} = d)$ for every $d \in \{0,1\}$ and $|\widehat{\lambda_0} - 1| = 0.44 > c_n$, we use partial identification results of Theorem 3.3 to estimate bounds for $F_{Y_{11}(0)|C}(y)$ and $F_{Y_{11}(1)|C}(y)$. The resulting estimates are displayed in Figure 3. The bounds are wide. The resulting bounds for QTE are uninformative, as 0 always lies between $\widehat{\underline{\tau}}_q$ and $\widehat{\overline{\tau}}_q$. Our IV-CIC model does not allow us to draw informative conclusions on returns to education from the natural experiment analyzed in Duflo (2001). This is because the increase in treatment rate was not much larger in high than in low treatment regions. As a result, $\widehat{\mu}_1 = 1.11$ is not much larger than $\widehat{\lambda}_1 = 1.05$, and $\widehat{\mu}_0 = 0.53$ is not much smaller than $\widehat{\lambda}_0 = 0.56$. In Appendix C, we study an application in which our bounds are still informative despite the fact that the treatment rate substantially increases in the control group, because $\widehat{\mu}_0$ is much smaller than $\widehat{\lambda}_0$.

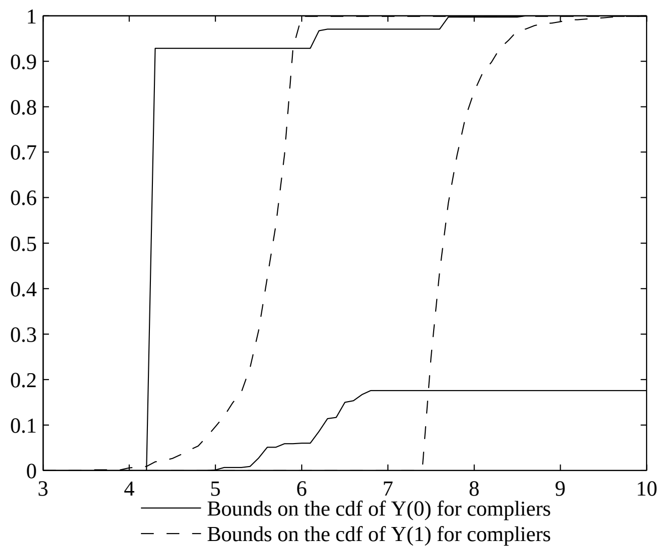

Figure 3: Estimated bounds on the cdf of $Y(0)$ and $Y(1)$ for compliers.

This application shows that when exposition to treatment substantially changes in the control group as well, using our IV-CIC model may result in wide and uninformative bounds. In such instances, point identification can still be achieved using IV-DID, but this strategy relies on more stringent conditions than our IV-CIC model (de Chaisemartin, 2013). Besides common trend conditions on potential outcomes, a common trend condition on potential treatments is also needed. Here, the share of individuals completing primary school increased by 4.5 percentage points in the low treatment regions. IV-DID then requires that the share of individuals completing primary school would have also increased by 4.5 percentage points in high treatment regions if as many schools had been constructed there than in low treatment regions. It also requires an homogeneity condition on the returns to education across the two groups of regions. This may not hold if for instance the returns to education are heterogeneous across local labor markets.

Finally, another strategy to recover point identification would be to look for another control group in which educational attainment did not change over time, and then use our IV-CIC model. One could for instance use regions in which primary school completion rate changed the least across the two cohorts.

7 Conclusion

In this paper, we develop an IV-CIC model to identify treatment effects when the treatment rate increases more in some groups than in others, for instance following a legislative change. Our model brings several improvements to IV-DID, the model currently used in the literature to identify treatment effects in such settings. It does not require common trend assumptions, it is invariant to monotonic transforms of the outcome, and it does not impose that some subgroups of observations in the treatment and in the control groups have the same treatment effects.

We show that when the treatment rate is stable between period 0 and 1 in the control group, a LATE and QTE among compliers are point identified under our IV-CIC assumptions. When the treatment rate also changes between period 0 and 1 in the control group, the same LATE and QTE are partially identified. The smaller the change in the treatment rate in the control group, the tighter the bounds. We conduct inference on treatment effects and sharp bounds estimators by proving their asymptotic normality and showing the validity of the bootstrap.

A Main proofs

Even though it appears after Theorems 3.1 and 3.2 in the text, we start by proving Lemma 3.1, as the two aforementioned theorems follow from this lemma.

Lemma 3.1

We only prove the formula for $d = 0$, the reasoning being similar for $d = 1$. We first show that

$$F_{Y_{11}(0)|C}(y) = \frac{P(D_{10} = 0)F_{Y_{11}(0)|V<v_0(0)}(y) - P(D_{11} = 0)F_{Y_{11}|D=0}(y)}{P(D_{10} = 0) - P(D_{11} = 0)}. \tag{6}$$

To this aim, note first that

$$\begin{aligned}
P(C|G=1,T=1,V<v_0(0)) &= \frac{P(V \in [v_1(1), v_0(0))|G=1,T=1)}{P(V<v_0(0)|G=1,T=1)} \\
&= \frac{P(V<v_0(0)|G=1,T=1) - P(V<v_1(1)|G=1,T=1)}{P(V<v_0(0)|G=1,T=1)} \\
&= \frac{P(V<v_0(0)|G=1,T=0) - P(V<v_1(1)|G=1,T=1)}{P(V<v_0(0)|G=1,T=0)} \\
&= \frac{P(D_{10}=0) - P(D_{11}=0)}{P(D_{10}=0)}.
\end{aligned}$$

The third equality stems from Assumption 3, and $P(D_{10} = 0) > 0$ because of Assumption 5. Then

$$\begin{aligned}
F_{Y_{11}(0)|V<v_0(0)}(y) &= P(V \in [v_1(1), v_0(0))|G=1,T=1,V<v_0(0))F_{Y_{11}(0)|V\in[v_1(1),v_0(0))}(y) \\
&\quad + P(V<v_1(1)|G=1,T=1,V<v_0(0))F_{Y_{11}|V<v_1(1)}(y) \\
&= \frac{P(D_{10}=0) - P(D_{11}=0)}{P(D_{10}=0)}F_{Y_{11}(0)|C}(y) + \frac{P(D_{11}=0)}{P(D_{10}=0)}F_{Y_{11}|D=0}(y)
\end{aligned}$$

This proves (6), and thus the second point of the lemma.

To prove the first point of the lemma, we show that for all $y \in \mathcal{S}(Y_{11}(0)|V < v_0(0))$,

$$F_{Y_{11}(0)|V<v_0(0)} = F_{Y_{10}|D=0} \circ F^{-1}_{Y_{00}|D=0} \circ F_{Y_{01}(0)|V<v_0(0)}. \tag{7}$$

By Assumption 3, $(U_0, \mathbb{1}\{V < v_0(0)\}) \perp\!\!\!\perp T|G$, which implies

$$U_0 \perp\!\!\!\perp T|G, V < v_0(0).$$

As a result, for all $(g,t) \in \{0,1\}^2$,

$$\begin{aligned}
F_{Y_{gt}(0)|V<v_0(0)}(y) &= P(h_0(U_0,t) \leq y|G=g,T=t,V<v_0(0)) \\
&= P(U_0 \leq h_0^{-1}(y,t)|G=g,T=t,V<v_0(0)) \\
&= P(U_0 \leq h_0^{-1}(y,t)|G=g,V<v_0(0)) \\
&= F_{U_0|G=g,V<v_0(0)}(h_0^{-1}(y,t)).
\end{aligned}$$

The second point of Assumption 4 combined with Assumptions 1 and 3 implies that $F_{U_0|G=g,V<v_0(0)}$ is strictly increasing. Hence, its inverse exists and for all $q \in (0,1)$,

$$F^{-1}_{Y_{gt}(0)|V<v_0(0)}(q) = h_0\left(F^{-1}_{U_0|G=g,V<v_0(0)}(q), t\right).$$

This implies that for all $y \in \mathcal{S}(Y_{g1}(0)|V < v_0(0))$,

$$F^{-1}_{Y_{g0}(0)|V<v_0(0)} \circ F_{Y_{g1}(0)|V<v_0(0)}(y) = h_0(h_0^{-1}(y,1),0), \tag{8}$$

which is independent of g.

Now, we have

$$\begin{aligned}
&\mathcal{S}(Y_{10}|D=0) = \mathcal{S}(Y_{00}|D=0)\\
&\Rightarrow \mathcal{S}(Y_{10}(0)|V < v_0(0)) = \mathcal{S}(Y_{00}(0)|V < v_0(0))\\
&\Rightarrow \mathcal{S}(h_0(U_0,0)|V < v_0(0), G=1, T=0) = \mathcal{S}(h_0(U_0,0)|V < v_0(0), G=0, T=0)\\
&\Rightarrow \mathcal{S}(U_0|V < v_0(0), G=1) = \mathcal{S}(U_0|V < v_0(0), G=0)\\
&\Rightarrow \mathcal{S}(h_0(U_0,1)|V < v_0(0), G=1, T=1) = \mathcal{S}(h_0(U_0,1)|V < v_0(0), G=0, T=1)\\
&\Rightarrow \mathcal{S}(Y_{11}(0)|V < v_0(0)) = \mathcal{S}(Y_{01}(0)|V < v_0(0)),
\end{aligned}$$

where the third and fourth implications are obtained combining Assumptions 1 and 3. Therefore, for all $y \in \mathcal{S}(Y_{11}(0)|V < v_0(0))$,

$$F^{-1}_{Y_{10}(0)|V<v_0(0)} \circ F_{Y_{11}(0)|V<v_0(0)}(y) = F^{-1}_{Y_{00}(0)|V<v_0(0)} \circ F_{Y_{01}(0)|V<v_0(0)}(y).$$

This proves (7), because $V < v_0(0)$ is equivalent to $D = 0$ when $T = 0$, and because the second point of Assumption 4 implies that $F^{-1}_{Y_{10}|D=0}$ is strictly increasing on $(0,1)$.

Finally, we show that

$$F_{Y_{01}(0)|V<v_0(0)}(y) = \lambda_0 F_{Y_{01}|D=0}(y) + (1-\lambda_0) F_{Y_{01}(0)|TC}(y). \tag{9}$$

Suppose first that $\lambda_0 \le 1$. Then, $v_0(1) \le v_0(0)$ and TC is equivalent to the event $V \in [v_0(1), v_0(0))$. Moreover, reasoning as for $P(C|G=1, V < v_0(0))$, we get

$$\lambda_0 = \frac{P(V < v_0(1)|G=0)}{P(V < v_0(0)|G=0)} = P(V < v_0(1)|G=0, V < v_0(0)).$$

Then

$$\begin{aligned}
F_{Y_{01}(0)|V<v_0(0)}(y) &= P(V < v_0(1)|G=0, V < v_0(0)) F_{Y_{01}(0)|V<v_0(1)}(y)\\
&\quad + P(V \in [v_0(1), v_0(0))|G=0, V < v_0(0)) F_{Y_{01}|V\in[v_0(1),v_0(0))}(y)\\
&= \lambda_0 F_{Y_{01}|D=0}(y) + (1-\lambda_0) F_{Y_{01}(0)|TC}(y).
\end{aligned}$$

If $\lambda_0 > 1$, $v_0(1) > v_0(0)$ and TC is equivalent to the event $V \in [v_0(0), v_0(1))$.

$$\frac{1}{\lambda_0} = P(V < v_0(0)|G = 0, V < v_0(1))$$

and

$$F_{Y_{01}|D=0}(y) = \frac{1}{\lambda_0} F_{Y_{01}(0)|V<v_0(0)}(y) + \left(1 - \frac{1}{\lambda_0}\right) F_{Y_{01}(0)|TC}(y),$$

so that we also get (9).

Finally, the first point of the lemma follows by combining (6), (7) and (9).

Theorem 3.1

The proof follows directly from Lemma 3.1, by noting that $\lambda_0 = \lambda_1 = 1$ when $P(D_{00} = d) = P(D_{01} = d) > 0$.

Theorem 3.2

Assume that $P(D_{00} = 0) = P(D_{01} = 0) = 0$ (the proof is symmetric when $P(D_{00} = 1) = P(D_{01} = 1) = 0$). This implies that $P(D_{00} = 1) = P(D_{01} = 1) > 0$, so for $F_{Y_{11}(1)|C}(y)$ the proof directly follows from Lemma 3.1, by noting that $\lambda_1 = 1$.

For $F_{Y_{11}(0)|C}(y)$, one can use the same steps as in the proof of Lemma 3.1 to show that Equation (6) also holds here:

$$F_{Y_{11}(0)|C}(y) = \frac{P(D_{10} = 0)F_{Y_{11}(0)|V<v_0(0)}(y) - P(D_{11} = 0)F_{Y_{11}|D=0}(y)}{P(D_{10} = 0) - P(D_{11} = 0)}. \tag{10}$$

Then, let $\underline{v}$ denote the lower bound of $\mathcal{S}(V|G = 0)$. Following similar steps as those used to establish Equation (8), one can show that for all $y \in \mathcal{S}(Y_{01}(0)|V < v_0(0)) = \mathcal{S}(Y_{00}(0)|V \geq v_0(0)) = \mathcal{S}(Y)$,

$$F^{-1}_{Y_{10}(0)|V<v_0(0)} \circ F_{Y_{11}(0)|V<v_0(0)}(y) = h_0(h_0^{-1}(y, 1), 0),$$
$$F^{-1}_{Y_{00}(1)|V\geq\underline{v}} \circ F_{Y_{01}(1)|V\geq\underline{v}}(y) = h_1(h_1^{-1}(y, 1), 0).$$

Under Assumption 6, this implies that for all $y \in \mathcal{S}(Y)$,

$$\begin{aligned} F_{Y_{11}(0)|V<v_0(0)}(y) &= F_{Y_{10}(0)|V<v_0(0)} \circ F^{-1}_{Y_{00}(1)|V\geq\underline{v}} \circ F_{Y_{01}(1)|V\geq\underline{v}}(y) \\ &= F_{Y_{10}|D=0} \circ F^{-1}_{Y_{00}|D=1} \circ F_{Y_{01}|D=1}(y), \end{aligned} \tag{11}$$

where the second equality follows from the fact that $P(D_{00} = 1) = P(D_{01} = 1) = 1$. Combining Equations (10) and (11) yields the result for $F_{Y_{11}(0)|C}(y)$.

Theorem 3.3

We focus on the case where $P(D_{00} = d) > 0$, since the proofs for the case $P(D_{00} = d) = 0$ are immediate.

1. Construction of the bounds.

We only establish the validity of the bounds for $d = 0$, the reasoning being similar for $d = 1$. We start by considering the case where $\lambda_0 < 1$. We first show that in such instances, $0 \leq T_0, G_0(T_0), C_0(T_0) \leq 1$ if and only if

$$\underline{T_0} \leq T_0 \leq \overline{T_0}. \tag{12}$$

Indeed, $G_0(T_0)$ is included between 0 and 1 if and only if

$$\frac{-\lambda_0 F_{Y_{01}|D=0}}{1-\lambda_0} \leq T_0 \leq \frac{1-\lambda_0 F_{Y_{01}|D=0}}{1-\lambda_0},$$

while $C_0(T_0)$ is included between 0 and 1 if and only if

$$\frac{H_0^{-1}(\mu_0 F_{Y_{11}|D=0}) - \lambda_0 F_{Y_{01}|D=0}}{1-\lambda_0} \leq T_0 \leq \frac{H_0^{-1}(\mu_0 F_{Y_{11}|D=0} + (1-\mu_0)) - \lambda_0 F_{Y_{01}|D=0}}{1-\lambda_0}.$$

Since $-\lambda_0 F_{Y_{01}|D=0}/(1-\lambda_0) \leq 0$ and $(1-\lambda_0 F_{Y_{01}|D=0})/(1-\lambda_0) \geq 1$, T_0, $G_0(T_0)$ and $C_0(T_0)$ are all included between 0 and 1 if and only if

$$M_0\left(\frac{H_0^{-1}(\mu_0 F_{Y_{11}|D=0}) - \lambda_0 F_{Y_{01}|D=0}}{1-\lambda_0}\right) \leq T_0 \leq m_1\left(\frac{H_0^{-1}(\mu_0 F_{Y_{11}|D=0} + (1-\mu_0)) - \lambda_0 F_{Y_{01}|D=0}}{1-\lambda_0}\right). \tag{13}$$

By composing each term of these inequalities by $M_0(.)$ and then by $m_1(.)$, we obtain (12) since $M_0(T_0) = m_1(T_0) = T_0$ and $M_0 \circ m_1 = m_1 \circ M_0$.

Now, when $\lambda_0 < 1$, $G_0(T_0)$ is increasing in T_0, so $C_0(T_0)$ as well is increasing in T_0. Combining this with (12) implies that for every y',

$$C_0(\underline{T_0})(y') \leq C_0(T_0)(y') \leq C_0(\overline{T_0})(y').$$

Because $C_0(T_0)(y)$ is a cdf,

$$C_0(T_0)(y) = \inf_{y' \geq y} C_0(T_0)(y') \leq \inf_{y' \geq y} C_0(\overline{T_0})(y'). \tag{14}$$

The lower bound follows similarly.

Let us now turn to the case where $\lambda_0 > 1$. Using the same reasoning as above, we get that $G_0(T_0)$ and $C_0(T_0)$ are included between 0 and 1 if and only if

$$\frac{\lambda_0 F_{Y_{01}|D=0} - 1}{\lambda_0 - 1} \leq T_0 \leq \frac{\lambda_0 F_{Y_{01}|D=0}}{\lambda_0 - 1},$$

$$\frac{\lambda_0 F_{Y_{01}|D=0} - H_0^{-1}(\mu_0 F_{Y_{11}|D=0} + (1-\mu_0))}{\lambda_0 - 1} \leq T_0 \leq \frac{\lambda_0 F_{Y_{01}|D=0} - H_0^{-1}(\mu_0 F_{Y_{11}|D=0})}{\lambda_0 - 1}.$$

The inequalities in the first line are not binding since they are implied by those on the second line. Thus, we also get (13). Hence, using the same argument as previously,

$$\overline{T_0} \leq T_0 \leq \underline{T_0}. \tag{15}$$

Besides, when $\lambda_0 > 1$, $G_0(T_0)$ is decreasing in T_0, so that $C_0(T_0)$ as well is decreasing in T_0. Combining this with (15) implies that for every y, (14) holds as well. This proves the result.

2. Sharpness.

We only consider the sharpness of $\underline{B}_0$, the reasoning being similar for the upper bound. The proof is also similar and actually simpler for $d = 1$. The corresponding bounds are indeed proper cdf, so that we do not have to consider converging sequences of cdf as we do in case b) below.

a. $\lambda_0 > 1$**.** We show that if Assumptions 4-7 hold, then $\underline{B}_0$ is sharp. For that purpose, we construct $\widetilde{h}_0, \widetilde{U}_0, \widetilde{V}$ such that:

(i) $Y = \widetilde{h}_0(\widetilde{U}_0, T)$ when $D = 0$ and $D = 1\{\widetilde{V} \geq v_Z(T)\}$;

(ii) $\widetilde{h}_0(., t)$ is strictly increasing for $t \in \{0, 1\}$;

(iii) $(\widetilde{U}_0, \widetilde{V}) \perp\!\!\!\perp T|G$;

(iv) $F_{\widetilde{h}_0(\widetilde{U}_0,1)|G=0,T=1,\widetilde{V}\in[v_0(0),v_0(1))} = \underline{T}_0$.

Point (i) ensures that Equation (1) and Assumption 2 are satisfied on the observed data. Because we can always define $\widetilde{Y}(0)$ as $\widetilde{h}_0(\widetilde{U}_0, T)$ when $D = 1$ and $\widetilde{D}(z) = 1\{\widetilde{V} \geq v_z(T)\}$ when $Z \neq z$ without contradicting the data and the model, Point (i) is actually sufficient for Equation (1) and Assumption 2 to hold globally, not only on observed data. Point (ii) and (iii) ensure that Assumptions 1 and 3 hold. Finally, Point (iv) ensures that the DGP corresponding to $(\widetilde{h}_0, \widetilde{U}_0, \widetilde{V})$ rationalizes the bound. If $(\widetilde{h}_0, \widetilde{U}_0, \widetilde{V})$ satisfy Assumptions 1-5 and are such that $\widetilde{T}_0 = \underline{T}_0$, we can apply Lemma 3.1 to show that the bound is attained.

First, let

$$\begin{aligned} \widetilde{h}_0(., 0) &= F^{-1}_{Y_{00}|D=0} \circ G_0(\underline{T}_0) \circ F^{-1}_{Y_{01}|D=0}, \\ \widetilde{h}_0(., 1) &= F^{-1}_{Y_{01}|D=0}. \end{aligned}$$

Second, let

$$\begin{aligned} \widetilde{U}_0 = &\ (1 - D)\widetilde{h}_0^{-1}(Y, T) \\ &+ D(1 - T)(1 - G)1\{V \in [v_0(0), v_0(1))\}\widetilde{U}_0^1 \\ &+ DTG1\{V \in [v_1(1), v_0(0))\}\widetilde{U}_0^2 \\ &+ D\left[1 - (1 - T)(1 - G)1\{V \in [v_0(0), v_0(1))\} - TG1\{V \in [v_1(1), v_0(0))\}\right] U_0, \end{aligned}$$

where $\widetilde{U}_0^1$ and $\widetilde{U}_0^2$ are two random variables such that $\mathcal{S}(\widetilde{U}_0^1) = \mathcal{S}(\widetilde{U}_0^2) = (0,1)$, and

$$F_{\widetilde{U}_0^1|G=0,T=0,V\in[v_0(0),v_0(1))} = \underline{T}_0 \circ F^{-1}_{Y_{01}|D=0},$$
$$F_{\widetilde{U}_0^2|G=1,T=1,V\in(v_1(1),v_0(0))} = C_0(\underline{T}_0) \circ F^{-1}_{Y_{01}|D=0}.$$

$F_{\widetilde{U}_0^1|G=0,T=0,V\in[v_0(0),v_0(1))}$ is a valid cdf on $(0,1)$ since (i) $\underline{T}_0$ is increasing by Assumption 7 and $F^{-1}_{Y_{01}|D=0}$ is also increasing, (ii) $\lim_{y\to\underline{y}} \underline{T}_0(y) = 0$ and $\lim_{y\to\overline{y}} \underline{T}_0(y) = 1$ when $\lambda_0 > 1$. $F_{\widetilde{U}_0^2|G=0,T=0,V\in[v_0(0),v_0(1))}$ is also a valid cdf on $(0,1)$ since (i) $C_0(\underline{T}_0)$ is increasing by Assumption 7 and $F^{-1}_{Y_{01}|D=0}$ is also increasing, (ii) $C_0(\underline{T}_0)(\mathcal{S}(Y)) = (0,1)$ when $\lambda_0 > 1$, as per the second point of Lemma B.1.

Third, for every $u \in (0,1)$, let

$$\begin{aligned} P_0(u) &= \underline{T}_0 \circ F^{-1}_{Y_{01}|D=0}(u), \\ P_1(u) &= C_0(\underline{T}_0) \circ F^{-1}_{Y_{01}|D=0}(u), \\ P_2(u) &= H_0 \circ G_0(\underline{T}_0) \circ F^{-1}_{Y_{01}|D=0}(u). \end{aligned}$$

As shown in the proof of Lemma B.5 (lower bound, case 2), Assumption 7 ensures that $P_0(u)$, $P_1(u)$, and $P_2(u)$ are non differentiable at only one point. Moreover, using the fact that

$$F_{Y_{01}|D=0} = \frac{1}{\lambda_0} G_0(\underline{T}_0) + \left(1 - \frac{1}{\lambda_0}\right) \underline{T}_0, \tag{16}$$
$$H_0 \circ G_0(\underline{T}_0) = \mu_0 F_{Y_{11}|D=0} + (1-\mu_0) C_0(\underline{T}_0), \tag{17}$$

and $\underline{T}_0$, $G(\underline{T}_0)$, and $C_0(\underline{T}_0)$ are increasing under Assumption 7, one can show that

$$0 \leq \left(1 - \frac{1}{\lambda_0}\right) P_0'(u) \leq 1,$$
$$0 \leq \frac{(1-\mu_0) P_1'(u)}{P_2'(u)} \leq 1,$$

for any u at which $P_0(.)$, $P_1(.)$, and $P_2(.)$ are differentiable, and $P_2'(u) > 0$. Then, let B_{TC} and B_C be two Bernoulli random variables such that for every $u \in (0,1)$,

$$P(B_{TC} = 1|\widetilde{U}_0 = u, D = 0, G = 0, T = 1) = \left(1 - \frac{1}{\lambda_0}\right) P_0'(u),$$
$$P(B_C = 1|\widetilde{U}_0 = u, D = 0, G = 1, T = 0) = \frac{(1-\mu_0) P_1'(u)}{P_2'(u)},$$

with the convention that $P(B_{TC} = 1|\widetilde{U}_0 = u, D = 0, G = 0, T = 1)$ and $P(B_C = 1|\widetilde{U}_0 = u, D = 0, G = 1, T = 0)$ are equal to 0 at the point at which $P_0(u)$, $P_1(u)$, and $P_2(u)$ are not differentiable, and $P(B_C = 1|\widetilde{U}_0 = u, D = 0, G = 1, T = 0) = 0$ when $P_2'(u) = 0$. The first convention is innocuous as it applies to a 0 Lebesgue measure set. As we shall see later, the second convention is also innocuous, because when $P_2'(u)=0$, Equation (17) implies that $P_1'(u)=0$ as well.

Finally, let

$$\begin{aligned}\widetilde{V} &= (1-D)(1-G)T\left[B_{TC}\widetilde{V}^1+(1-B_{TC})\widetilde{V}^2\right]\\ &+(1-D)G(1-T)\left[B_C\widetilde{V}^3+(1-B_C)\widetilde{V}^4\right]\\ &+(1-(1-D)\left[(1-G)T+G(1-T)\right])V,\end{aligned}$$

where $\widetilde{V}^1$, $\widetilde{V}^2$, $\widetilde{V}^3$, and $\widetilde{V}^4$ are such that $\mathcal{S}(\widetilde{V}^1)=\mathcal{S}(V)\cap[v_0(0),v_0(1))$, $\mathcal{S}(\widetilde{V}^2)=\mathcal{S}(V)\cap(-\infty,v_0(0))$, $\mathcal{S}(\widetilde{V}^3)=\mathcal{S}(V)\cap(v_1(1),v_0(0))$, $\mathcal{S}(\widetilde{V}^4)=\mathcal{S}(V)\cap(-\infty,v_1(1))$, and

$$\begin{aligned}f_{\widetilde{V}^1|G=0,T=1,D=0,B_{TC}=1,\widetilde{U}_0}(v|u) &= f_{V|G=0,T=0,V\in[v_0(0),v_0(1)),\widetilde{U}_0}(v|u),\\ f_{\widetilde{V}^2|G=0,T=1,D=0,B_{TC}=0,\widetilde{U}_0}(v|u) &= f_{V|G=0,T=0,V<v_0(0),\widetilde{U}_0}(v|u),\\ f_{\widetilde{V}^3|G=1,T=0,D=0,B_C=1,\widetilde{U}_0}(v|u) &= f_{V|G=1,T=1,V\in[v_1(1),v_0(0)),\widetilde{U}_0}(v|u),\\ f_{\widetilde{V}^4|G=1,T=0,D=0,B_C=0,\widetilde{U}_0}(v|u) &= f_{V|G=1,T=1,V<v_1(1),\widetilde{U}_0}(v|u).\end{aligned}$$

We shall now show that $(\widetilde{h}_0(.,0),\widetilde{h}_0(.,1),\widetilde{U}_0,\widetilde{V})$ satisfies (i), (ii), (iii), and (iv). By construction, Point (i) is satisfied. Moreover, it follows from Assumption 4 that $\widetilde{h}_0(.,1)$ is strictly increasing on $(0,1)$. Besides, $G_0(\underline{T}_0)\circ F^{-1}_{Y_{01}|D=0}$ is strictly increasing on $(0,1)$ and included between 0 and 1 as shown in the first point of Lemma B.1. $F^{-1}_{Y_{00}|D=0}$ is also strictly increasing on $(0,1)$ by Assumption 4. Therefore, $\widetilde{h}_0(.,0)$ is also strictly increasing on $(0,1)$, and Point (ii) is satisfied.

Then, we check Point (iii). We show that it holds in the control group. For that purpose, we use Bayes law to write

$$\begin{aligned}&f_{\widetilde{U}_0,\widetilde{V}|G=0,T=t}(u,v)\\ =\ &P(\widetilde{V}<v_0(1)|G=0,T=t)[P(\widetilde{V}<v_0(0)|G=0,T=t,\widetilde{V}<v_0(1))f_{\widetilde{U}_0|G=0,T=t,\widetilde{V}<v_0(0)}(u)f_{\widetilde{V}|G=0,T=t,\widetilde{V}<v_0(0),\widetilde{U}_0}(v|u)\\ &+P(\widetilde{V}\in[v_0(0),v_0(1))|G=0,T=t,\widetilde{V}<v_0(1))f_{\widetilde{U}_0|G=0,T=t,\widetilde{V}\in[v_0(0),v_0(1))}(u)f_{\widetilde{V}|G=0,T=t,\widetilde{V}\in[v_0(0),v_0(1)),\widetilde{U}_0}(v|u)]\\ &+P(\widetilde{V}\geq v_0(1)|G=0,T=t)f_{\widetilde{U}_0,\widetilde{V}|G=0,T=t,\widetilde{V}\geq v_0(1)}(u,v),\end{aligned}\tag{18}$$

and we show that all elements in the right-hand side of the previous display are equal for $t=0$ and $t=1$.

We first evaluate all of these quantities when $T=1$. First, it follows from the definition of $\widetilde{V}$ that

$$P(\widetilde{V}<v_0(1)|G=0,T=1)=P(D_{01}=0).\tag{19}$$

Then,

$$\begin{aligned}
P(\widetilde{U}_0 \leq u|G=0,T=1,\widetilde{V}<v_0(1)) &= P(\widetilde{U}_0 \leq u|G=0,T=1,D=0) \\
&= P(\widetilde{h}_0^{-1}(Y,1) \leq u|G=0,T=1,D=0) \\
&= P(Y \leq F^{-1}_{Y_{01}|D=0}(u)|G=0,T=1,D=0) \\
&= u.
\end{aligned}$$

Therefore,

$$f_{\widetilde{U}_0|G=0,T=1,\widetilde{V}<v_0(1)}(u) = 1.$$

Then, we have, almost everywhere,

$$\begin{aligned}
& f_{\widetilde{U}_0,1\{\widetilde{V}\in[v_0(0),v_0(1))\}|G=0,T=1,\widetilde{V}<v_0(1)}(u,1) \\
= \ & P(\widetilde{V}\in[v_0(0),v_0(1))|G=0,T=1,\widetilde{V}<v_0(1),\widetilde{U}_0=u)f_{\widetilde{U}_0|G=0,T=1,\widetilde{V}<v_0(1)}(u) \\
= \ & P(B_{TC}=1|G=0,T=1,D=0,\widetilde{U}_0=u) \\
= \ & \left(1-\frac{1}{\lambda_0}\right)P_0'(u). \qquad (20)
\end{aligned}$$

The second equality follows from the definition of $\widetilde{V}$, and from $f_{\widetilde{U}_0|G=0,T=1,\widetilde{V}<v_0(1)}(u) = 1$. Equation (20) and the fact that P_0' is a density imply that

$$P(\widetilde{V}\in[v_0(0),v_0(1))|G=0,T=1,\widetilde{V}<v_0(1)) = 1-\frac{1}{\lambda_0}, \qquad (21)$$

$$f_{\widetilde{U}_0|G=0,T=1,\widetilde{V}\in[v_0(0),v_0(1))}(u) = P_0'(u), \qquad (22)$$

and

$$P(\widetilde{V}<v_0(0)|G=0,T=1,\widetilde{V}<v_0(1)) = \frac{1}{\lambda_0}, \qquad (23)$$

$$f_{\widetilde{U}_0|G=0,T=1,\widetilde{V}<v_0(0)}(u) = \lambda_0-(\lambda_0-1)\,P_0'(u). \qquad (24)$$

Next, we have

$$\begin{aligned}
f_{\widetilde{V}|G=0,T=1,\widetilde{V}\in[v_0(0),v_0(1)),\widetilde{U}_0}(v|u) &= f_{\widetilde{V}^1|G=0,T=1,D=0,B_{TC}=1,\widetilde{U}_0}(v|u), \\
&= f_{V|G=0,T=0,V\in[v_0(0),v_0(1)),\widetilde{U}_0}(v|u), \qquad (25)
\end{aligned}$$

and

$$\begin{aligned}
f_{\widetilde{V}|G=0,T=1,\widetilde{V}<v_0(0),\widetilde{U}_0}(v|u) &= f_{\widetilde{V}^2|G=0,T=1,D=0,B_{TC}=0,\widetilde{U}_0}(v|u) \\
&= f_{V|G=0,T=0,V<v_0(0),\widetilde{U}_0}(v|u). \qquad (26)
\end{aligned}$$

Then, we evaluate all of these quantities when $T=0$. First, notice that

$$\begin{aligned}
P(\widetilde{V}<v_0(1)|G=0,T=0) &= P(V<v_0(1)|G=0,T=0) \\
&= P(V<v_0(1)|G=0,T=1) \\
&= P(D_{01}=0). \qquad (27)
\end{aligned}$$

The first equality follows from the definition of $\widetilde{V}$ and the second from the fact V satisfies Assumption 3. One can use similar arguments to show that

$$P(\widetilde{V} \in [v_0(0), v_0(1))|G = 0, T = 0, \widetilde{V} < v_0(1)) = 1 - \frac{1}{\lambda_0}, \quad (28)$$

$$P(\widetilde{V} < v_0(0)|G = 0, T = 0, \widetilde{V} < v_0(1)) = \frac{1}{\lambda_0}. \quad (29)$$

Then, it follows from the definition of $\widetilde{V}$ and $\widetilde{U}_0$ that

$$f_{\widetilde{U}_0|G=0,T=0,\widetilde{V}\in[v_0(0),v_0(1))}(u) = f_{\widetilde{U}_0^1|G=0,T=0,V\in[v_0(0),v_0(1))}(u) = P_0'(u). \quad (30)$$

Next,

$$\begin{aligned} P(\widetilde{U}_0 \leq u|G = 0, T = 0, \widetilde{V} < v_0(0)) &= P(\widetilde{U}_0 \leq u|G = 0, T = 0, D = 0) \\ &= P(\widetilde{h}_0^{-1}(Y, 0) \leq u|G = 0, T = 0, D = 0) \\ &= P(Y \leq F^{-1}_{Y_{00}|D=0} \circ G_0(\underline{T}_0) \circ F^{-1}_{Y_{01}|D=0}(u)|G = 0, T = 0, D = 0) \\ &= G_0(\underline{T}_0) \circ F^{-1}_{Y_{01}|D=0}(u) \\ &= \lambda_0 u - (\lambda_0 - 1)\, P_0(u), \end{aligned}$$

where the last equality follows from (16). This implies that

$$f_{\widetilde{U}_0|G=0,T=0,\widetilde{V}<v_0(0)}(u) = \lambda_0 - (\lambda_0 - 1)\, P_0'(u). \quad (31)$$

Then, it follows from the definition of $\widetilde{V}$ that

$$f_{\widetilde{V}|G=0,T=0,\widetilde{V}\in[v_0(0),v_0(1)),\widetilde{U}_0}(v|u) = f_{V|G=0,T=0,V\in[v_0(0),v_0(1)),\widetilde{U}_0}(v|u), \quad (32)$$

$$f_{\widetilde{V}|G=0,T=0,\widetilde{V}<v_0(0),\widetilde{U}_0}(v|u) = f_{V|G=0,T=0,V<v_0(0),\widetilde{U}_0}(v|u). \quad (33)$$

Finally,

$$\begin{aligned} f_{\widetilde{U}_0,\widetilde{V}|G=0,T=0,\widetilde{V}\geq v_0(1)}(u, v) &= f_{U_0,V|G=0,T=0,V\geq v_0(1)}(u, v) \\ &= f_{U_0,V|G=0,T=1,V\geq v_0(1)}(u, v) \\ &= f_{\widetilde{U}_0,\widetilde{V}|G=0,T=1,\widetilde{V}\geq v_0(1)}(u, v), \quad (34) \end{aligned}$$

where the first and last equality follow from the definition of $(\widetilde{U}_0, \widetilde{V})$, while the second equality follows from the fact (U_0, V) satisfies Assumption 3.

Finally, combining Equation (18) with Equations (19) and (27), (21) and (28), (23) and (29), (22) and (30), (24) and (31), (25) and (32), (26) and (33), and (34), we get that

$$f_{\widetilde{U}_0,\widetilde{V}|G=0,T=1}(u, v) = f_{\widetilde{U}_0,\widetilde{V}|G=0,T=0}(u, v).$$

This shows that (iii) holds in the control group. Showing that it also holds in the treatment group relies on a very similar reasoning, so we skip this part of the proof due to a concern for brevity.

b. $\lambda_0 < 1$**.** The idea is similar as in the previous case. A difference, however, is that when $\lambda_0 < 1$, $\underline{T}_0$ is not a proper cdf, but a defective one, since $\lim_{y\to\overline{y}} \underline{T}_0(y) < 1$. As a result, we cannot define a DGP such that $\widetilde{T}_0 = \underline{T}_0$, However, by Lemma B.2, there exists a sequence $(\underline{T}_0^k)_k$ of cdf such that $\underline{T}_0^k \to \underline{T}_0$, $G_0(\underline{T}_0^k)$ is an increasing bijection from $\mathcal{S}(Y)$ to $(0,1)$ and $C_0(\underline{T}_0^k)$ is increasing and onto $(0,1)$. We can then define a sequence of DGP $(\widetilde{h}_0^k(.,0), \widetilde{h}_0^k(.,1), \widetilde{U}_0^k, \widetilde{V}^k)$ such that Points (i) to (iii) listed above hold for every k, and such that $\widetilde{T}_0^k = \underline{T}_0^k$. Since $\underline{T}_0^k(y)$ converges to $\underline{T}_0(y)$ for every y in $\overset{\circ}{\mathcal{S}(Y)}$, we thus define a sequence of DGP such that $\widetilde{T}_0^k$ can be arbitrarily close to $\underline{T}_0$ on $\overset{\circ}{\mathcal{S}(Y)}$ for sufficiently large k. Since $C_0(.)$ is continuous, this proves that $\underline{B}_0$ is sharp on $\overset{\circ}{\mathcal{S}(Y)}$.

In what follows, we exhibit $\widetilde{h}_0^k(.,0)$ and $\widetilde{h}_0^k(.,1)$ satisfying (i), as well as distributions of $\widetilde{U}_0^k$ for all relevant subpopulations which are a) compatible with the data, b) satisfy (iii), and c) reach the bound. We do not not exhibit $(\widetilde{U}_0^k, \widetilde{V}^k)$ as we did in the previous proof, to avoid repeating twice similar arguments.

Let

$$\begin{aligned} \widetilde{h}_0^k(.,1) &= G_0(\underline{T}_0^k)^{-1} \\ \widetilde{h}_0^k(.,0) &= F^{-1}_{Y_{00}|D=0} \end{aligned}$$

$\widetilde{h}_0^k(.,1)$ is strictly increasing on $(0,1)$ since $G_0(\underline{T}_0^k)$ is an increasing bijection on $(0,1)$ as shown in Lemma B.2. $\widetilde{h}_0^k(.,0)$ is strictly increasing on $(0,1)$ under Assumption 4. Therefore, (i) is verified.

Let us consider first the distribution of $\widetilde{U}_0^k$ among untreated observations in the control group in period 1. It follows from Bayes rule that

$$F_{\widetilde{U}_0^k|G=0,T=1,\widetilde{V}<v_0(0)} = \lambda_0 F_{\widetilde{U}_0^k|G=0,T=1,\widetilde{V}<v_0(1)} + (1-\lambda_0) F_{\widetilde{U}_0^k|G=0,T=1,\widetilde{V}\in[v_0(1),v_0(0))} \qquad (35)$$

Given $\widetilde{h}_0^k(.,1)$, to have $\widetilde{T}_0^k = \underline{T}_0^k$, we must have

$$F_{\widetilde{U}_0^k|G=0,T=1,\widetilde{V}\in[v_0(1),v_0(0))} = \underline{T}_0^k \circ G_0(\underline{T}_0^k)^{-1}.$$

This defines a valid cdf since $\underline{T}_0^k$ is a cdf and $G_0(\underline{T}_0^k)^{-1}$ is increasing and onto $\mathcal{S}(Y)$. It can be achieved by constructing $\widetilde{V}$ using an appropriate Bernoulli random variable to split untreated observations in the control group in period 0 between some for which $\widetilde{V} \in [v_0(1), v_0(0))$, and some for which $\widetilde{V} < v_0(1)$, exactly as we did for $\lambda_0 > 1$.

Given $\widetilde{h}_0^k(.,1)$, and the fact $\widetilde{h}_0^k(\widetilde{U}_0^k,1)=Y$ for all observations such that $G=0,T=1,\widetilde{V}<v_0(1)$, a few computations yield

$$F_{\widetilde{U}_0^k|G=0,T=1,\widetilde{V}<v_0(1)}=F_{Y_{01}|D=0}\circ G_0(\underline{T}_0^k)^{-1}.$$

Plugging the last two equations into (35) finally yields $F_{\widetilde{U}_0^k|G=0,T=1,\widetilde{V}<v_0(0)}=I$, where I denotes the identity function on $[0,1]$.

Now, let us turn to untreated observations in the control group in period 0. Given $\widetilde{h}_0^k(.,0)$, and the fact $\widetilde{h}_0^k(\widetilde{U}_0^k,0)=Y$ for all observations such that $G=0,T=0,\widetilde{V}<v_0(0)$, a few computations yield $F_{\widetilde{U}_0^k|G=0,T=0,\widetilde{V}<v_0(0)}=I$. Since $Y(0)$ is not observed for observations such that $G=0,T=1,\widetilde{V}\in[v_0(1),v_0(0))$, the data does not impose any constraint on their U_0, so we can set

$$F_{\widetilde{U}_0^k|G=0,T=0,\widetilde{V}\in[v_0(1),v_0(0))}=\underline{T}_0^k\circ G_0(\underline{T}_0^k)^{-1}.$$

Therefore, the distributions of $\widetilde{U}_0^k|G=0,T=t,\widetilde{V}<v_0(1)$ and $\widetilde{U}_0^k|G=0,T=t,\widetilde{V}\in[v_0(1),v_0(0))$ satisfy (iii).

Then, let us consider untreated observations in the treatment group in period 1. Using the definition of $\widetilde{h}_0^k(.,1)$ and the fact $\widetilde{h}_0^k(\widetilde{U}_0^k,1)=Y$ for all observations such that $G=1,T=1,\widetilde{V}<v_1(1)$, one can show after a few computations that

$$F_{\widetilde{U}_0^k|G=1,T=1,\widetilde{V}<v_1(1)}=F_{Y_{11}|D=0}\circ G_0(\underline{T}_0^k)^{-1}.$$

Since $Y(0)$ is not observed for observations such that $G=1,T=1,\widetilde{V}\in[v_1(1),v_0(0))$, the data does not impose any constraint on their U_0, so we can set

$$F_{\widetilde{U}_0^k|G=1,T=1,\widetilde{V}\in[v_1(1),v_0(0))}=C_0(\underline{T}_0^k)\circ G_0(\underline{T}_0^k)^{-1}.$$

This defines a valid cdf, as shown in Points 2 and 3 of Lemma B.2.

Finally, let us consider untreated observations in the treatment group in period 0. It follows from Bayes rule that we must have

$$F_{\widetilde{U}_0^k|G=1,T=0,\widetilde{V}<v_0(0)}=\mu_0 F_{\widetilde{U}_0^k|G=1,T=0,\widetilde{V}<v_1(1)}+(1-\mu_0)F_{\widetilde{U}_0^k|G=1,T=0,\widetilde{V}\in[v_1(1),v_0(0))}. \quad (36)$$

To satisfy point (iii), we must have

$$F_{\widetilde{U}_0^k|G=1,T=0,\widetilde{V}<v_1(1)}=F_{Y_{11}|D=0}\circ G_0(\underline{T}_0^k)^{-1}.$$

This can be achieved by constructing $\widetilde{V}$ using an appropriate Bernoulli random variable to split untreated observations in the treatment group in period 0 between some for which $\widetilde{V}\in[v_1(1),v_0(0))$, and some for which $\widetilde{V}<v_1(1)$, exactly as we did for $\lambda_0>1$. Using the definition of $\widetilde{h}_0^k(.,1)$ and the fact $\widetilde{h}_0^k(\widetilde{U}_0^k,1)=Y$ for all observations such that $G=0,T=1,\widetilde{V}<v_1(1)$, one can show after a few computations that

$$F_{\widetilde{U}_0^k|G=1,T=0,\widetilde{V}<v_0(0)} = F_{Y_{10}|D=0}\circ F^{-1}_{Y_{00}|D=0}.$$

Plugging the last two equations into (36) finally yields

$$\begin{aligned} F_{\widetilde{U}_0^k|G=1,T=0,\widetilde{V}\in[v_1(1),v_0(0))} &= \frac{P(D_{10}=0)F_{Y_{10}|D=0}\circ F^{-1}_{Y_{00}|D=0} - P(D_{11}=0)F_{Y_{11}|D=0}\circ G_0(\underline{T}_0^k)^{-1}}{P(D_{10}=0)-P(D_{11}=0)} \\ &= C_0(\underline{T}_0^k)\circ G_0(\underline{T}_0^k)^{-1}. \end{aligned}$$

Therefore, the distributions of $\widetilde{U}_0^k|G=1,T=t,\widetilde{V}<v_1(1)$ and $\widetilde{U}_0^k|G=1,T=t,\widetilde{V}\in[v_1(1),v_0(0))$ satisfy (iii). This completes the proof when $\lambda_0<1$.

Corollary 3.4

The bounds on Δ and τ_q are a direct consequence of Theorem 3.3. Note that the bounds on the LATE are well defined under Assumption 8. We now prove that these bounds are sharp under Assumption 7. We only focus on the lower bound, the result being similar for the upper bound. The model and data impose no condition on the joint distribution of (U_0,U_1). Hence, by the proof Theorem 3.3, we can rationalize the fact that $(F_{Y_{11}(0)|C},F_{Y_{11}(1)|C})=(\underline{B}_0,\overline{B}_1)$ when $\lambda_0>1$. Sharpness of Δ and τ_q follows directly. When $\lambda_0<1$, on the other hand, we can only rationalize the fact that $(F_{Y_{11}(0)|C},F_{Y_{11}(1)|C})=(G_{0k},F_{Y_{11}(1)|C})$, where G_{0k} converges pointwise to $\underline{B}_0$. To show the sharpness of the LATE and QTE, we thus have to prove that $\lim_{k\to\infty}\int ydG_{0k}(y)=\int yd\underline{B}_0(y)$ and $\lim_{k\to\infty}G_{0k}^{-1}(q)=\underline{B}_0^{-1}(q)$.

As for the LATE, we have, by integration by parts for Lebesgue-Stieljes integrals,

$$\int ydG_{0k}(y)=\overline{y}-\int_{\underline{y}}^{\overline{y}}G_{0k}dy=-\int_{\underline{y}}^{0}G_{0k}(y)dy+\int_0^{\overline{y}}\left[1-G_{0k}(y)\right]dy. \tag{37}$$

We now prove the convergence of each integral in the right-hand side. As shown by Lemma B.2, G_{0k} can be defined as $G_{0k}=C_0(\underline{T}_0^k)$ with $\underline{T}_0^k\leq T_0$, T_0 denoting the true cdf of $Y_{11}(0)$ for time compliers, which satisfies $C_0(T_0)=F_{Y_{11}(0)|C}$. Because $C_0(.)$ is increasing when $\lambda_0<1$, $G_{0k}\leq F_{Y_{11}(0)|C}$. $E(|Y_{11}(0)|\ |C)<+\infty$ implies that $\int_{\underline{y}}^0 F_{Y_{11}(0)|C}(y)dy<+\infty$. Thus, by the dominated convergence theorem,

$$\lim_{k\to\infty}\int_{\underline{y}}^0 G_{0k}dy=\int_{\underline{y}}^0\underline{B}_0(y)dy<+\infty.$$

Now consider the second integral in (37). If $\overline{y}<+\infty$, we can also apply the dominated convergence theorem: $1-G_{0k}\leq 1$ implies that $\int_0^{\overline{y}}\left[1-G_{0k}(y)\right]dy\to\int_0^{\overline{y}}\left[1-\underline{B}_0(y)\right]dy$. If

$\overline{y} = +\infty$, $\lim_{y \to +\infty} \underline{B}_0(y) = \ell < 1$ so that

$$\int_0^{\overline{y}} [1 - \underline{B}_0(y)]\, dy = +\infty.$$

By Fatou's lemma,

$$\liminf \int_0^{\overline{y}} [1 - G_{0k}(y)]\, dy \geq \int_0^{\overline{y}} [1 - \underline{B}_0(y)]\, dy = +\infty.$$

Thus, in this case as well the second integral in (37) converges to $\int_0^{\overline{y}} [1 - \underline{B}_0(y)]\, dy$. Finally, because $\int_{\underline{y}}^0 G_{0k}(y)dy$ converges to a finite limit, $\int y dG_{0k}(y)$ converges to $\int y d\underline{B}_0(y)$. Hence, the lower bound of Δ is sharp.

Now, let us turn to τ_q. Let y_0 be the unique solution to $\underline{B}_0(y) = q$. Fix $\varepsilon > 0$. By pointwise convergence, Assumption 7 and unicity of the solution of $\underline{B}_0(y) = q$,

$$\lim_{k \to \infty} G_{0k}(y_0 + \varepsilon) = \underline{B}_0(y_0 + \varepsilon) > \underline{B}_0(y_0) = q.$$

Thus, there exists k_0 such that for all $k \geq k_0$, $G_{0k}(y_0 + \varepsilon) > q$. As a result, by definition of G_{0k}^{-1}, $y_0 + \varepsilon > G_{0k}^{-1}(q)$ for all $k \geq k_0$. Similarly, there exists k_1 such that for all $k \geq k_1$, $y_0 - \varepsilon < G_{0k}^{-1}(q)$. Hence, for all $k \geq \max(k_0, k_1)$, $|G_{0k}^{-1}(q) - \overline{B}_0^{-1}(q)| < \varepsilon$. The result follows.

Lemma 4.1

First, use Bayes' rule to write:

$$f_{X_{11}|C}(x) = \frac{P(C|X_{11} = x) f_{X_{11}}(x)}{P(C)}.$$

Then, notice that

$$\begin{aligned}
P(C|X_{11} = x) &= P(V \in [v_1(1,x), v_0(0,x))|X_{11} = x) \\
&= P(V \geq v_1(1,x)|X_{11} = x) - P(V \geq v_0(0,x)|X_{11} = x) \\
&= P(V \geq v_1(1,x)|X = x, G = 1, T = 1) - P(V \geq v_0(0,x)|X = x, G = 1, T = 1) \\
&= P(V \geq v_1(1,x)|X = x, G = 1, T = 1) - P(V \geq v_0(0,x)|X = x, G = 1, T = 0) \\
&= P(D = 1|X = x, G = 1, T = 1) - P(D = 1|X = x, G = 1, T = 0) \\
&= P(D_{11} = 1|X = x) - P(D_{10} = 1|X = x).
\end{aligned}$$

The third step follows from Assumption 11. Therefore,

$$P(C) = E\left[P(D_{11} = 1|X) - P(D_{10} = 1|X)|G = 1, T = 1\right].$$

The result follows combining the three previous equalities.

Theorem 4.2

We only prove the result for $d = 0$, the reasoning being similar for $d = 1$.

Because we always have $D(0) = D$ in the control group, Assumption 12 implies that for every $x \in \mathcal{S}(X)$ and for $t \in \{0, 1\}$, $F_{Y_{0t}(0)|D(0)=0,X=x}$ is strictly increasing, and $\mathcal{S}(Y_{0t}(0)|D(0) = 0, X = x) = \mathcal{S}(Y_{0t}|D = 0, X = x) = \mathcal{S}(Y)$. Then, let $h_0^{-1}(., t, x)$ denote the inverse of $h_0(., t, x)$. For every $(x, y) \in \mathcal{S}(X) \times \mathcal{S}(Y)$,

$$\begin{aligned} F_{Y_{0t}(0)|D(0)=0,X=x}(y) &= P(h_0(U_0, t, x) \leq y|G = 0, T = t, D(0) = 0, X = x) \\ &= P(U_0 \leq h_0^{-1}(y, t, x)|G = 0, T = t, D(0) = 0, X = x) \\ &= P(U_0 \leq h_0^{-1}(y, t, x)|G = 0, D(0) = 0, X = x) \\ &= F_{U_0|G=0,D(0)=0,X=x}(h_0^{-1}(y, t, x)), \end{aligned}$$

where the first equality stems from Assumption 9 and the third from Assumption 14. Therefore, we also have that for all $q \in (0, 1)$,

$$F^{-1}_{Y_{0t}(0)|D(0)=0,X=x}(q) = h_0\left(F^{-1}_{U_0|G=0,D(0)=0,X=x}(q), t, x\right).$$

This implies that for every $(x, y) \in \mathcal{S}(X) \times \mathcal{S}(Y)$,

$$F^{-1}_{Y_{00}(0)|D(0)=0,X=x} \circ F_{Y_{01}(0)|D(0)=0,X=x}(y) = h_0(h_0^{-1}(y, 1, x), 0, x). \tag{38}$$

Then, notice that

$$\begin{aligned} \mathcal{S}(Y_{11}(0)|V < v_0(0, x), X = x) &= \mathcal{S}(h_0(U_0, 1, x)|G = 1, T = 1, V < v_0(0, x), X = x) \\ &= \mathcal{S}(h_0(U_0, 1, x)|G = 1, T = 0, V < v_0(0, x), X = x) \\ &= \mathcal{S}(h_0(U_0, 1, x)|G = 0, T = 0, V < v_0(0, x), X = x) \\ &= \mathcal{S}(h_0(U_0, 1, x)|G = 0, T = 0, D(0) = 0, X = x) \\ &= \mathcal{S}(h_0(U_0, 1, x)|G = 0, T = 1, D(0) = 0, X = x) \\ &= \mathcal{S}(Y_{01}|D = 0, X) \\ &= \mathcal{S}(Y). \end{aligned}$$

The second equality follows from Assumption 11. The third follows from Assumption 12: $\mathcal{S}(Y_{10}|D = 0, X = x) = \mathcal{S}(Y_{00}|D = 0, X = x)$ implies that

$$\mathcal{S}(U_0|G = 1, T = 0, V < v_0(0, x), X = x) = \mathcal{S}(U_0|G = 0, T = 0, V < v_0(0, x), X = x).$$

The fifth follows from Assumption 14.

As we also have

$$\mathcal{S}(Y_{10}(0)|V < v_0(0,x), X = x) = \mathcal{S}(Y_{10}|D = 0, X = x) = \mathcal{S}(Y),$$

one can use Assumption 11 to prove a conditional version of Equation (8): for every $(x, y) \in \mathcal{S}(X) \times \mathcal{S}(Y)$,

$$F^{-1}_{Y_{10}(0)|V<v_0(0,x),X=x} \circ F_{Y_{11}(0)|V<v_0(0,x),X=x}(y) = h_0(h_0^{-1}(y, 1, x), 0, x). \tag{39}$$

Combining Equations (38) and (39) implies that for every $(x, y) \in \mathcal{S}(X) \times \mathcal{S}(Y)$,

$$F_{Y_{11}(0)|V<v_0(0,x),X=x}(y) = F_{Y_{10}|D=0,X=x} \circ F^{-1}_{Y_{00}|D=0,X=x} \circ F_{Y_{01}|D=0,X=x}(y). \tag{40}$$

Finally, under our conditional IV-CIC model we can prove a conditional version of Equation (6):

$$F_{Y_{11}(0)|C,X=x}(y) = \frac{P(D_{10} = 0|X = x)F_{Y_{11}(0)|V<v_0(0,x),X=x}(y) - P(D_{11} = 0|X = x)F_{Y_{11}|D=0,X=x}(y)}{P(D_{10} = 0|X = x) - P(D_{11} = 0|X = x)}.$$

Plugging (40) into the last equation yields the result.

Theorem 4.3

We prove the two points by contradiction. In each case we focus on $d = 0$, the proof being similar for $d = 1$.

Point 1. Let us assume that $\mathcal{T}_0 \neq \emptyset$. Then there exists a function T_0 increasing and included in $[0, 1]$ such that $G_0(T_0)$ and $C_0(T_0)$ are also increasing and included in $[0, 1]$. As shown in (12), when $\lambda_0 \leq 1$, $0 \leq T_0, G_0(T_0), C_0(T_0) \leq 1$ implies that we must have

$$\underline{T_0} \leq T_0 \leq \overline{T_0}.$$

Conversely, as shown in (15), when $\lambda_0 > 1$, $0 \leq T_0, G_0(T_0), C_0(T_0) \leq 1$ implies

$$\overline{T_0} \leq T_0 \leq \underline{T_0}.$$

Therefore, $0 \leq T_0, G_0(T_0), C_0(T_0) \leq 1$ always implies

$$\min(\overline{T_0}, \underline{T_0}) \leq T_0 \leq \max(\overline{T_0}, \underline{T_0}). \tag{41}$$

Moreover, T_0 increasing implies

$$T_0(y_0) \leq T_0(y_1). \tag{42}$$

Combining Equations (41) and (42) implies that we must have

$$\min(\overline{T_0}(y_0), \underline{T_0}(y_0)) \leq T_0(y_0) \leq T_0(y_1) \leq \max(\overline{T_0}(y_1), \underline{T_0}(y_1)). \tag{43}$$

This contradicts the fact that $I_0(y_0, y_1) = \emptyset$. Hence, $\mathcal{T}_0 = \emptyset$, implying that we reject Assumptions 1-3 together.

Point 2. Now assume that there exists $y_0 < y_1$ in $\mathcal{S}(Y)^2$ such that $I_0(y_0, y_1) \neq \emptyset$ and for every $t_0 \leq t_1$ in $I_0(y_0, y_1)^2$,

$$\begin{aligned} & \frac{P(D_{10}=0)H_0 \circ (\lambda_0 F_{Y_{01}|D=0}(y_1) + (1-\lambda_0)t_1) - P(D_{11}=0)F_{Y_{11}|D=0}(y_1)}{P(D_{10}=0) - P(D_{11}=0)} \\ < \ & \frac{P(D_{10}=0)H_0 \circ (\lambda_0 F_{Y_{01}|D=0}(y_0) + (1-\lambda_0)t_0) - P(D_{11}=0)F_{Y_{11}|D=0}(y_0)}{P(D_{10}=0) - P(D_{11}=0)}. \end{aligned} \tag{44}$$

Assume also that $\mathcal{T}_0 \neq \emptyset$. $C_0(T_0)$ increasing implies

$$\begin{aligned} & \frac{P(D_{10}=0)H_0 \circ (\lambda_0 F_{Y_{01}|D=0}(y_1) + (1-\lambda_0)T_0(y_1)) - P(D_{11}=0)F_{Y_{11}|D=0}(y_1)}{P(D_{10}=0) - P(D_{11}=0)} \\ \geq \ & \frac{P(D_{10}=0)H_0 \circ (\lambda_0 F_{Y_{01}|D=0}(y_0) + (1-\lambda_0)T_0(y_0)) - P(D_{11}=0)F_{Y_{11}|D=0}(y_0)}{P(D_{10}=0) - P(D_{11}=0)}. \end{aligned} \tag{45}$$

As shown above in Equation (43), the fact that we must have $0 \leq T_0, G_0(T_0), C_0(T_0) \leq 1$ and T_0 increasing implies that we must have $T_0(y_0) \leq T_0(y_1)$ and $(T_0(y_0), T_0(y_1)) \in I_0(y_0, y_1)^2$, which is not empty by assumption. Combining this with Equation (45) proves that there exists $t_0 = T_0(y_0) \leq t_1 = T_0(y_1)$ in $I_0(y_0, y_1)^2$ such that

$$\begin{aligned} & \frac{P(D_{10}=0)H_0 \circ (\lambda_0 F_{Y_{01}|D=0}(y_1) + (1-\lambda_0)t_1) - P(D_{11}=0)F_{Y_{11}|D=0}(y_1)}{P(D_{10}=0) - P(D_{11}=0)} \\ \geq \ & \frac{P(D_{10}=0)H_0 \circ (\lambda_0 F_{Y_{01}|D=0}(y_0) + (1-\lambda_0)t_0) - P(D_{11}=0)F_{Y_{11}|D=0}(y_0)}{P(D_{10}=0) - P(D_{11}=0)}. \end{aligned}$$

This contradicts (44). Hence $\mathcal{T}_0 = \emptyset$, and once more we reject Assumptions 1-3 together.

Theorem 5.1

Hereafter, we let $\mathcal{C}^0$ and $\mathcal{C}^1$ denote respectively the set of continuous functions and the set of continuously differentiable functions with strictly positive derivative on $\mathcal{S}(Y)$.

We first show that $(\widehat{F}_{Y_{11}(0)|C}, \widehat{F}_{Y_{11}(1)|C})$ tends to a continuous gaussian process. Let $\widetilde{\theta} = (F_{000}, F_{001}, ..., F_{111}, \mu_0, \mu_1)$. By Lemma B.3, $\widehat{\widetilde{\theta}} = (\widehat{F}_{000}, \widehat{F}_{001}, ..., \widehat{F}_{111}, \widehat{\mu}_0, \widehat{\mu}_1)$ converges to a continuous gaussian process. Let

$$\pi_d : (F_{000}, F_{001}, ..., F_{111}, \mu_0, \mu_1) \mapsto (F_{d10}, F_{d00}, F_{d01}, F_{d11}, 1, \mu_d), \quad d \in \{0, 1\},$$

so that $(\widehat{F}_{Y_{11}(0)|C}, \widehat{F}_{Y_{11}(1)|C}) = \left(R_1 \circ \pi_0(\widetilde{\theta}), R_1 \circ \pi_1(\widetilde{\theta})\right)$, where R_1 is defined as in Lemma B.4. π_d is Hadamard differentiable as a linear continuous map. Because $F_{d10}, F_{d00}, F_{d01}, F_{d11}$ are continuously differentiable with strictly positive derivative by Assumption 16, $\mu_d > 0$, and $\mu_d \neq 1$ under Assumption 4, R_1 is also Hadamard differentiable at $(F_{d10}, F_{d00}, F_{d01}, F_{d11}, 1, \mu_d)$ tangentially to $(\mathcal{C}^0)^4 \times \mathbb{R}$. By the functional delta method (see, e.g., van der Vaart & Wellner, 1996, Lemma 3.9.4), $(\widehat{F}_{Y_{11}(0)|C}, \widehat{F}_{Y_{11}(1)|C})$ tends to a continuous gaussian process.

Now, by integration by parts for Lebesgue-Stieljes integrals,

$$\Delta = \int_{\underline{y}}^{\overline{y}} F_{Y_{11}(0)|C}(y) - F_{Y_{11}(1)|C}(y)dy.$$

Moreover, the map $\varphi_1 : (F_1, F_2) \mapsto \int_{\mathcal{S}(Y)}(F_2(y) - F_1(y))dy$, defined on the domain of bounded càdlàg functions, is linear. Because $\mathcal{S}(Y)$ is bounded by Assumption 16, φ_1 is also continuous with respect to the supremum norm. It is thus Hadamard differentiable. Because $\widehat{\Delta} = \varphi_1\left(\widehat{F}_{Y_{11}(1)|C}, \widehat{F}_{Y_{11}(0)|C}\right)$, $\widehat{\Delta}$ is asymptotically normal by the functional delta method. The asymptotic normality of $\widehat{\tau}_q$ follows along similar lines. By Assumption 16, $F_{Y_{11}(d)|C}$ is differentiable with strictly positive derivative on its support. Thus, the map $(F_1, F_2) \mapsto F_2^{-1}(q) - F_1^{-1}(q)$ is Hadamard differentiable at $(F_{Y_{11}(0)|C}, F_{Y_{11}(1)|C})$ tangentially to the set of functions that are continuous at $(F^{-1}_{Y_{11}(0)|C}(q), F^{-1}_{Y_{11}(1)|C}(q))$ (see Lemma 21.3 in van der Vaart, 2000). By the functional delta method, $\widehat{\tau}_q$ is asymptotically normal.

The validity of the bootstrap follows along the same lines. By Lemma B.3, the bootstrap is consistent for $\widehat{\theta}$. Because both the LATE and QTE are Hadamard differentiable functions of $\widehat{\theta}$, as shown above, the result simply follows by the functional delta method for the bootstrap (see, e.g., van der Vaart, 2000, Theorem 23.9).

Theorem 5.2

Let $\theta = (F_{000}, ..., F_{011}, F_{100}, ..., F_{111}, \lambda_0, \mu_0, \lambda_1, \mu_1)$. By Lemma B.5, for $d \in \{0, 1\}$ and $q \in \mathcal{Q}$, $\theta \mapsto \int_{\underline{y}}^{\overline{y}} \underline{B}_d(y)dy$, $\theta \mapsto \int_{\underline{y}}^{\overline{y}} \overline{B}_d(y)dy$, $\theta \mapsto \overline{B}_d^{-1}(q)$, and $\theta \mapsto \underline{B}_d^{-1}(q)$ are Hadamard differentiable tangentially to $(\mathcal{C}^0)^4 \times \mathbb{R}^2$. Because $\underline{\Delta} = \int_{\mathcal{S}(Y)} \underline{B}_0(y) - \overline{B}_1(y)dy$, $\underline{\Delta}$ is also a Hadamard differentiable function of θ tangentially to $(\mathcal{C}^0)^4 \times \mathbb{R}^2$. The same reasoning applies for $\overline{\Delta}$, and for $\underline{\tau_q}$ and $\overline{\tau_q}$ for every $q \in \mathcal{Q}$. The theorem then follows from Lemma B.3, the functional delta method, and the functional delta method for the bootstrap.

B Technical lemmas

Lemma B.1 *Assume Assumptions 4 and 7 hold, and $\lambda_d > 1$. Then:*

1. *$G_d(\underline{T}_d)$ is a bijection from $\mathcal{S}(Y)$ to $[0,1]$;*

2. *$C_d(\underline{T}_d)(\mathcal{S}(Y)) = [0,1]$.*

Proof: we only prove the result for $d = 0$, the reasoning being similar otherwise. One can show that when $\lambda_0 > 1$,

$$G_0(\underline{T}_0) = \min\left(\lambda_0 F_{Y_{01}|D=0}, \max\left(\lambda_0 F_{Y_{01}|D=0} + (1-\lambda_0), H_0^{-1} \circ \left(\mu_0 F_{Y_{11}|D=0}\right)\right)\right). \tag{46}$$

By Assumption 4, $\mu_0 F_{Y_{11}|D=0}$ is strictly increasing. Moreover, $\mathcal{S}(Y_{10}|D=0) = \mathcal{S}(Y_{00}|D=0)$ implies that H_0^{-1} is strictly increasing on $[0,1]$. Consequently, $H_0^{-1} \circ \left(\mu_0 F_{Y_{11}|D=0}\right)$ is strictly increasing on $\mathcal{S}(Y)$ since $\mu_0 < 1$. Therefore, $G_0(\underline{T}_0)$ is strictly increasing on $\mathcal{S}(Y)$ as a composition of the max and min of strictly increasing functions, which in turn implies that $G_0(\underline{T}_0) \circ F^{-1}_{Y_{01}|D=0}$ is strictly increasing on $[0,1]$. Moreover, it is easy to see that since $\mathcal{S}(Y_{1t}|D=0) = \mathcal{S}(Y_{0t}|D=0)$,

$$\begin{aligned}
\lim_{y \to \underline{y}} H_0^{-1} \circ \left(\mu_0 F_{Y_{11}|D=0}\right) \circ F^{-1}_{Y_{01}|D=0}(y) &= 0, \\
\lim_{y \to \overline{y}} H_0^{-1} \circ \left(\mu_0 F_{Y_{11}|D=0}\right) \circ F^{-1}_{Y_{01}|D=0}(y) &\leq 1.
\end{aligned}$$

Hence, by Equation (46),

$$\lim_{y \to \underline{y}} G_0(\underline{T}_0)(y) = 0, \ \lim_{y \to \overline{y}} G_0(\underline{T}_0)(y) = 1. \tag{47}$$

Finally, $G_0(\underline{T}_0) \circ F^{-1}_{Y_{01}|D=0}$ is also continuous by Assumption 4, as a composition of continuous functions. Point 1 then follows, by the intermediate value theorem.

Now, we have

$$C_0(\underline{T}_0) = \frac{P(D_{10}=0) F_{Y_{10}|D=0} \circ F^{-1}_{Y_{01}|D=0} \circ G_0(\underline{T}_0) - P(D_{11}=0) F_{Y_{11}|D=0}}{P(D_{10}=0) - P(D_{11}=0)}.$$

(47) implies that $G_0(\underline{T}_0)$ is a cdf. Hence, by Assumption 4,

$$\lim_{y \to \underline{y}} C_0(\underline{T}_0)(y) = 0, \ \lim_{y \to \overline{y}} C_0(\underline{T}_0)(y) = 1.$$

Moreover, $C_0(\underline{T}_0)$ is increasing by Assumption 7. Combining this with Assumption 4 yields Point 2, since $C_0(\underline{T}_0)$ is continuous by Assumption 4 once more.

Lemma B.2 *Assume Assumptions 4 and 7 hold, $P(D_{g0}=0)>0$ for $g\in\{0;1\}$ and $\lambda_0<1$. Then there exists a sequence of cdf $\underline{T}_0^k$ such that*

1. $\underline{T}_0^k(y)\to \underline{T}_0(y)$ *for all* $y\in \overset{\circ}{\mathcal{S}(Y)}$;
2. $G_0(\underline{T}_0^k)$ *is an increasing bijection from* $\mathcal{S}(Y)$ *to* $[0,1]$;
3. $C_0(\underline{T}_0^k)$ *is increasing and onto* $[0,1]$.

The same holds for the upper bound.

Proof: we consider a sequence $(y_k)_{k\in\mathbb{N}}$ converging to $\overline{y}$ and such that $y_k<\overline{y}$. Since $y_k<\overline{y}$, we can also define a strictly positive sequence $(\eta_k)_{k\in\mathbb{N}}$ such that $y_k+\eta_k<\overline{y}$. By Assumption 7, H_0 is continuously differentiable. Moreover,

$$H_0'=\frac{F'_{Y_{10}|D=0}\circ F'_{Y_{00}|D=0}}{F'_{Y_{00}|D=0}\circ F'_{Y_{00}|D=0}}$$

is strictly positive on $\mathcal{S}(Y)$ under Assumption 7. $F'_{Y_{11}|D=0}$ is also strictly positive on $\mathcal{S}(Y)$ under Assumption 7. Therefore, using a Taylor expansion of H_0 and $F_{Y_{11}|D=0}$, it is easy to show that there exists constants $A_{1k}>0$ and $A_{2k}>0$ such that for all $y<y'\in[y_k,y_k+\eta_k]^2$,

$$\begin{aligned} H_0(y')-H_0(y) &\geq A_{1k}(y'-y), && (48)\\ F_{Y_{11}|D=0}(y')-F_{Y_{11}|D=0}(y) &\leq A_{2k}(y'-y). && (49)\end{aligned}$$

We also define a sequence $(\varepsilon_k)_{k\in\mathbb{N}}$ by

$$\varepsilon_k=\min\left(\eta_k,\frac{A_{1k}(1-\lambda_0)\left(T_0(y_k)-\underline{T}_0(y_k)\right)}{\mu_0A_{2k}}\right). \qquad (50)$$

Note that as shown in (12), since $\lambda_0<1$, $0\leq T_0,G_0(T_0),C_0(T_0)\leq 1$ implies that we must have

$$\underline{T}_0\leq T_0,$$

which implies in turn that $\varepsilon_k\geq 0$. Consequently, since $0\leq\varepsilon_k\leq\eta_k$, inequalities (48) and (49) also hold for $y<y'\in[y_k,y_k+\varepsilon_k]^2$.

We now define $\underline{T}_0^k$. For every k such that $\varepsilon_k>0$, let

$$\underline{T}_0^k(y)=\left|\begin{array}{ll}\underline{T}_0(y) & \text{if } y<y_k\\ \underline{T}_0(y_k)+\frac{T_0(y_k+\varepsilon_k)-\underline{T}_0(y_k)}{\varepsilon_k}(y-y_k) & \text{if } y\in[y_k,y_k+\varepsilon_k]\\ T_0(y) & \text{if } y>y_k+\varepsilon_k.\end{array}\right.$$

For every k such that $\varepsilon_k=0$, let

$$\underline{T}_0^k(y)=\left|\begin{array}{ll}\underline{T}_0(y) & \text{if } y<y_k\\ T_0(y) & \text{if } y\geq y_k\end{array}\right.$$

Then, we verify that $\underline{T}_0^k$ defines a sequence of cdf which satisfy Points 1, 2 and 3. Under Assumption 7, $\underline{T}_0(y)$ is increasing, which implies that $\underline{T}_0^k(y)$ is increasing on $(\underline{y}, y_k)$. Since $T_0(y)$ is a cdf, $\underline{T}_0^k(y)$ is also increasing on $(y_k+\varepsilon_k, \overline{y})$. Finally, it is easy to check that when $\varepsilon_k > 0$, $\underline{T}_0^k(y)$ is increasing on $[y_k, y_k+\varepsilon_k]$. $\underline{T}_0^k$ is continuous on $(\underline{y}, y_k)$ and $(y_k+\varepsilon_k, \overline{y})$ under Assumption 4. It is also continuous at y_k and $y_k+\varepsilon_k$ by construction. This proves that $\underline{T}_0^k(y)$ is increasing on $\mathcal{S}(Y)$. Moreover,

$$\begin{aligned}\lim_{y\to\underline{y}} \underline{T}_0^k(y) &= \lim_{y\to\underline{y}} \underline{T}_0(y) = 0,\\ \lim_{y\to\overline{y}} \underline{T}_0^k(y) &= \lim_{y\to\overline{y}} T_0(y) = 1.\end{aligned}$$

Hence, $\underline{T}_0^k$ is a cdf. Point 1 also holds by construction of $\underline{T}_0^k(y)$.

$G_0(\underline{T}_0^k) = \lambda_0 F_{Y_{01}|D=0} + (1-\lambda_0)\underline{T}_0^k$ is strictly increasing and continuous as a sum of the strictly increasing and continuous function $\lambda_0 F_{Y_{01}|D=0}$ and an increasing and continuous function. Moreover, $G_0(\underline{T}_0^k)$ tends to 0 (resp. 1) when y tends to $\underline{y}$ (resp. to $\overline{y}$). Point 2 follows by the intermediate value theorem.

Finally, let us show Point 3. Because $G_0(\underline{T}_0^k)$ is a continuous cdf, $C_0(\underline{T}_0^k)$ is also continuous and converges to 0 (resp. 1) when y tends to $\underline{y}$ (resp. to $\overline{y}$). Thus, the proof will be completed if we show that $C_0(\underline{T}_0^k)$ is increasing. By Assumption 7, $C_0(\underline{T}_0^k)$ is increasing on $(\underline{y}, y_k)$. Moreover, since $F_{Y_{11}(0)|C} = C_0(T_0)$, $C_0(\underline{T}_0^k)$ is also increasing on $(y_k+\varepsilon_k, \overline{y})$. Finally, when $\varepsilon_k > 0$, we have that for all $y < y' \in [y_k, y_k+\varepsilon_k]^2$,

$$\begin{aligned}& H_0(\lambda_0 F_{Y_{01}|D=0}(y') + (1-\lambda_0)\underline{T}_0^k(y')) - H_0(\lambda_0 F_{Y_{01}|D=0}(y) + (1-\lambda_0)\underline{T}_0^k(y))\\ \geq\ & A_{1k}(1-\lambda_0)\left(\underline{T}_0^k(y') - \underline{T}_0^k(y)\right)\\ \geq\ & \frac{A_{1k}(1-\lambda_0)\left(T_0(y_k) - \underline{T}_0(y_k)\right)}{\varepsilon_k}(y'-y)\\ \geq\ & \mu_0 A_{2k}(y'-y)\\ \geq\ & \mu_0\left(F_{Y_{11}|D=0}(y') - F_{Y_{11}|D=0}(y)\right),\end{aligned}$$

where the first inequality follows by (48) and $F_{Y_{01}|D=0}(y') \geq F_{Y_{01}|D=0}(y)$, the second by the definition of $\underline{T}_0^k$ and $T_0(y_k+\varepsilon_k) \geq T_0(y_k)$, the third by (50) and the fourth by (49). This implies that $C_0(\underline{T}_0^k)$ is increasing on $[y_k, y_k+\varepsilon_k]$, since

$$C_0(\underline{T}_0^k) = \frac{H_0(\lambda_0 F_{Y_{01}|D=0} + (1-\lambda_0)\underline{T}_0^k) - \mu_0 F_{Y_{11}|D=0}}{1-\mu_0}.$$

It is easy to check that under Assumption 4 $C_0(\underline{T}_0^k)$ is continuous on $\mathcal{S}(Y)$. This completes the proof.

Lemma B.3 *Suppose that $P(D_{g0} = d) > 0$ for $(d, g) \in \{0,1\}^2$ and let*

$$\theta = (F_{000}, F_{001}, ..., F_{111}, \lambda_0, \mu_0, \lambda_1, \mu_1)$$

and

$$\widehat{\theta} = (\widehat{F}_{000}, \widehat{F}_{001}, ..., \widehat{F}_{111}, \widehat{\lambda}_0, \widehat{\mu}_0, \widehat{\lambda}_1, \widehat{\mu}_1).$$

Then

$$\sqrt{n}\left(\widehat{\theta} - \theta\right) \Longrightarrow G,$$

where G denotes a continuous gaussian process. Moreover, the bootstrap is consistent for $\widehat{\theta}$.

Proof: let $\mathbb{G}_n$ denote the standard empirical process and $p_{dgt} = P(D_{gt} = d)$. We prove the result for $\eta = (F_{000}, F_{001}, ..., F_{111}, p_{000}, ..., p_{011})$. The result follows since $(\lambda_0, \mu_0, \lambda_1, \mu_1)$ is a smooth function of $(p_{000}, ..., p_{011})$. For any $(y, d, g, t) \in (\mathcal{S}(Y) \cup \{+\infty\}) \times \{0, 1\}^3$, let

$$f_{y,d,g,t}(Y, D, G, T) = \frac{\mathbb{1}\{D = d\}\mathbb{1}\{G = g\}\mathbb{1}\{T = t\}\mathbb{1}\{Y \leq y\}}{p_{dgt}}.$$

We have, for all $(y, d, g, t) \in (\mathcal{S}(Y) \cup \{-\infty, +\infty\}) \times \{0, 1\}^3$,

$$\begin{aligned}
\sqrt{n}\left(\widehat{F}_{dgt}(y) - F_{dgt}(y)\right) &= \frac{\sqrt{n}}{n_{dgt}} \sum_{i=1}^{n} \mathbb{1}\{D_i = d\}\mathbb{1}\{G_i = g\}\mathbb{1}\{T_i = t\}\mathbb{1}\{Y_i \leq y\} - F_{dgt}(y) \\
&= \frac{\sqrt{n}}{n_{dgt}} \sum_{i=1}^{n} \mathbb{1}\{D_i = d\}\mathbb{1}\{G_i = g\}\mathbb{1}\{T_i = t\} \left[\mathbb{1}\{Y_i \leq y\} - F_{dgt}(y)\right]. \\
&= \frac{np_{dgt}}{n_{dgt}} \mathbb{G}_n f_{y,d,g,t}.
\end{aligned}$$

Now, $\mathcal{F} = \left\{f_{y,d,g,t} : (y, d, g, t) \in (\mathcal{S}(Y) \cup \{+\infty\}) \times \{0, 1\}^3\right\}$ is Donsker (see, e.g., van der Vaart, 2000, Example 19.6). Besides, $np_{dgt}/n_{dgt} \xrightarrow{\mathbb{P}} 1$. Thus, by Slutski's lemma (see, e.g., van der Vaart, 2000, Theorem 18.10 (v)), $\widehat{\eta}$, the empirical counterpart of η, converges to a gaussian process.

Now let us turn to the bootstrap. Observe that

$$\sqrt{n}\left(\widehat{F}^*_{dgt}(y) - F_{dgt}(y)\right) = \frac{np_{dgt}}{n^*_{dgt}} \mathbb{G}^*_n f_{y,d,g,t},$$

where $\mathbb{G}^*_n$ denote the bootstrap empirical process. Because $np_{dgt}/n^*_{dgt} \xrightarrow{\mathbb{P}} 1$ and by consistency of the bootstrap empirical process (see, e.g., van der Vaart, 2000, Theorem 23.7), the bootstrap is consistent for $\widehat{\eta}$.

Notation for the next lemmas

In the following lemmas, we let, for any functional R, dR denote the Hadamard differential of R whenever it exists, that is to say the continuous linear form satisfying

$$dR(h) = \lim_{t \to 0} \frac{R(F + th_t) - R(F)}{t}, \text{ for any } h_t \text{ s.t. } ||h_t - h||_\infty \to 0.$$

Note that the point at which the differential is taken is left implicit.

Lemma B.4 *1. Let* $R_1(F_1, F_2, F_3, F_4, \lambda, \mu) = \frac{\mu F_4 - F_1 \circ F_2^{-1} \circ q_1(F_3,\lambda)}{\mu - 1}$ *and* $R_2(F_1, F_2, F_3, F_4, \lambda, \mu) = \frac{\mu F_4 - F_1 \circ F_2^{-1} \circ q_2(F_3,\lambda)}{\mu - 1}$*, with* $q_1(F_3, \lambda) = \lambda F_3$ *and* $q_2(F_3, \lambda) = \lambda F_3 + 1 - \lambda$*.* R_1 *and* R_2 *are Hadamard differentiable at any* $(F_{10}, F_{20}, F_{30}, F_{40}, \lambda_0, \mu_0) \in (\mathcal{C}^1)^4 \times [0, \infty) \times ([0, \infty) \backslash \{1\})$*, tangentially to* $(\mathcal{C}^0)^4 \times \mathbb{R}^2$*. Moreover,* $dR_1\left((\mathcal{C}^0)^4 \times \mathbb{R}^2\right)$ *and* $dR_2\left((\mathcal{C}^0)^4 \times \mathbb{R}^2\right)$ *are included in* $\mathcal{C}^0$*.*

2. Let $R_3(F_1) = \int_{\underline{y}}^{\overline{y}} m_1(F_1)(y)dy$ *and* $R_4(F_1, F_2) = \int_{\underline{y}}^{\overline{y}} F_2(m_1(F_1))(y)dy$*.* R_3 *is Hadamard differentiable at any* F_{10} *such that* F_{10} *is increasing on* $\mathcal{S}(Y)$ *and the equation* $F_{10}(y) = 1$ *admits at most one solution on* $\overset{\circ}{\mathcal{S}}(Y)$*, tangentially to* $\mathcal{C}^0$*.* R_4 *is Hadamard differentiable at any* (F_{10}, F_{20}) *such that* F_{10} *satisfies the same conditions as for* R_3 *and* F_{20} *is continuously differentiable on* $[0, 1]$*, tangentially to* $(\mathcal{C}^0)^2$*. The same holds if we replace* m_1 *(and the equation* $F_{10}(y) = 1$*) by* M_0 *(and* $F_{10}(y) = 0$*).*

Proof of 1. We first prove that $\phi_1(F_1, F_2, F_3) = F_1 \circ F_2^{-1} \circ F_3$ is Hadamard differentiable at $(F_{10}, F_{20}, F_{30}) \in \left(\mathcal{C}^1\right)^3$. Let $\mathcal{D}$ denote the set of bounded càdlàg functions on $[\underline{y}, \overline{y}]$. Because $(F_{10}, F_{20}) \in \left(\mathcal{C}^1\right)^2$, the function $\phi_2 : (F_1, F_2, F_3) \mapsto (F_1 \circ F_2^{-1}, F_3)$ is Hadamard differentiable at (F_{10}, F_{20}, F_{30}) tangentially to $\mathcal{D} \times \mathcal{C}^0 \times \mathcal{D}$ (see, e.g., van der Vaart & Wellner, 1996, Problem 3.9.4), and therefore tangentially to $\left(\mathcal{C}^0\right)^3$. Moreover computations show that its derivative at (F_{10}, F_{20}, F_{30}) satisfies

$$d\phi_2(h_1, h_2, h_3) = (h_1 \circ F_{20}^{-1} - \frac{F_{10}' \circ F_{20}^{-1}}{F_{20}' \circ F_{20}^{-1}} h_2 \circ F_{20}^{-1}, h_3).$$

This shows that $d\phi_2\left(\left(\mathcal{C}^0\right)^3\right) \subseteq \left(\mathcal{C}^0\right)^2$.

Then, the composition function $\phi_3 : (U, V) \mapsto U \circ V$ is Hadamard differentiable at any $(U_0, V_0) \in \left(\mathcal{C}^1\right)^2$, tangentially to $\mathcal{C}^0 \times \mathcal{D}$ (see, e.g., van der Vaart & Wellner, 1996, Lemma 3.9.27), and therefore tangentially to $\left(\mathcal{C}^0\right)^2$. It is thus Hadamard differentiable at $(F_{10} \circ F_{20}^{-1}, F_{30})$, and one can show that $d\phi_3\left(\left(\mathcal{C}^0\right)^2\right) \subseteq \mathcal{C}^0$. Thus, by the chain rule (see van der Vaart & Wellner, 1996, Lemma 3.9.3), $\phi_1 = \phi_3 \circ \phi_2$ is also Hadamard differentiable at (F_{10}, F_{20}, F_{30}) tangentially to $(\mathcal{C}^0)^3$, and $d\phi_1\left(\left(\mathcal{C}^0\right)^3\right) \subseteq \mathcal{C}^0$.

Finally, because $q_1(F_3, \lambda)$ is a smooth function of F_3 and λ, and R_1 is a smooth function of $(\phi_1(F_1, F_2, q_1(F_3, \lambda)), F_4, \mu)$, it is also Hadamard differentiable at $(F_{10}, F_{20}, F_{30}, F_{40}, \lambda_0, \mu_0)$ tangentially to $(\mathcal{C}^0)^4 \times \mathbb{R}^2$, and $dR_1\left((\mathcal{C}^0)^4 \times \mathbb{R}^2\right) \subseteq \mathcal{C}^0$.

Proof of 2. We only prove the result for R_4 and m_1, the reasoning being similar (and more simple) for R_3 and M_0. For any collections of functions (h_{t1}) and (h_{t2}) in $\mathcal{C}^0$, respectively converging uniformly towards h_1 and h_2 in $\mathcal{C}^0$, we have

$$\begin{aligned}\frac{R_4(F_{10}+th_{t1},F_{20}+th_{t2})-R_4(F_{10},F_{20})}{t} &= \int_{\underline{y}}^{\overline{y}} h_{t2}\circ m_1(F_{10}+th_{t1})(y)dy \\ &+ \int_{\underline{y}}^{\overline{y}} \frac{F_{20}\circ m_1(F_{10}+th_{t1})-F_{20}\circ m_1(F_{10})}{t}(y)dy.\end{aligned}$$

Consider the first integral I_1.

$$\begin{aligned}& |h_{t2}\circ m_1(F_{10}+th_{t1})(y)-h_2\circ m_1(F_{10})(y)| \\ \leq\ & |h_{t2}\circ m_1(F_{10}+th_{t1})(y)-h_2\circ m_1(F_{10}+th_{t1})(y)| \\ +\ & |h_2\circ m_1(F_{10}+th_{t1})(y)-h_2\circ m_1(F_{10})(y)| \\ \leq\ & ||h_{t2}-h_2||_\infty+|h_2\circ m_1(F_{10}+th_{t1})(y)-h_2\circ m_1(F_{10})(y)|.\end{aligned}$$

By uniform convergence of h_{t2} towards h_2, the first term in the last inequality converges to 0 when t goes to 0. By convergence of $m_1(F_{10}+th_{t1})$ towards $m_1(F_{10})$ and continuity of h_2, the second term also converges to 0. As a result,

$$h_{t2}\circ m_1(F_{10}+th_{t1})(y)\to h_2\circ m_1(F_{10})(y).$$

Moreover, for t small enough,

$$|h_{t2}\circ m_1(F_{10}+th_{t1})(y)|\leq ||h_2||_\infty+1.$$

Thus, by the dominated convergence theorem, $I_1\to\int_{\underline{y}}^{\overline{y}} h_2\circ m_1(F_{10})(y)dy$, which is linear in h_2 and continuous since the integral is taken over a bounded interval.

Now consider the second integral I_2. Let us define $\underline{y}_1$ as the solution to $F_{10}(y)=1$ on $(\underline{y},\overline{y})$ if there is one such solution, $\underline{y}_1=\overline{y}$ otherwise. We prove that almost everywhere,

$$\frac{F_{20}\circ m_1(F_{10}(y)+th_{t1}(y))-F_{20}\circ m_1(F_{10}(y))}{t}\to F_{20}'(F_{10}(y))h_1(y)\mathbb{1}\{y<\underline{y}_1\}. \qquad (51)$$

As F_{10} is increasing, for $y<\underline{y}_1$, $F_{10}(y)<1$, so that for t small enough, $F_{10}(y)+th_{t1}(y)<1$. Therefore, for t small enough,

$$\begin{aligned}\frac{F_{20}\circ m_1(F_{10}(y)+th_{t1}(y))-F_{20}\circ m_1(F_{10}(y))}{t} &= \frac{F_{20}\circ(F_{10}(y)+th_{t1}(y))-F_{20}\circ F_{10}(y)}{t} \\ &= \frac{(F_{20}'(F_{10}(y))+\varepsilon(t))(F_{10}(y)+th_{t1}(y)-F_{10}(y))}{t} \\ &= (F_{20}'(F_{10}(y))+\varepsilon(t))h_{t1}(y)\end{aligned}$$

for some function $\varepsilon(t)$ converging towards 0 when t goes to 0. Therefore,

$$\frac{F_{20}\circ m_1(F_{10}(y)+th_{t1}(y))-F_{20}\circ m_1(F_{10}(y))}{t}\to F_{20}'(F_{10}(y))h_1(y),$$

so that (51) holds for $y < \underline{y}_1$. Now, if $\overline{y} > y > \underline{y}_1$, $F_{10}(y) > 1$ because F_{10} is increasing. Thus, for t small enough, $F_{10}(y) + th_{t1}(y) > 1$. Therefore, for t small enough,

$$\frac{F_{20} \circ m_1(F_{10}(y) + th_{t1}(y)) - F_{20} \circ m_1(F_{10}(y))}{t} = 0,$$

so that (51) holds as well. Thus, (51) holds almost everywhere.

Now, remark that m_1 is 1-Lipschitz. As a result,

$$\begin{aligned} \left| \frac{F_{20} \circ m_1(F_{10}(y) + th_{t1}(y)) - F_{20} \circ m_1(F_{10}(y))}{t} \right| &\leq ||F'_{20}||_\infty |h_{t1}(y)| \\ &\leq ||F'_{20}||_\infty \left(|h_1(y)| + ||h_{t1} - h_1||_\infty\right). \end{aligned}$$

Because $||h_{t1} - h_1||_\infty \to 0$, $|h_1(y)| + ||h_{t1} - h_1||_\infty \leq |h_1(y)| + 1$ for t small enough. Thus, by the dominated convergence theorem,

$$\int_{\underline{y}}^{\overline{y}} \frac{F_{20} \circ m_1(F_{10} + th_{t1}) - F_{20} \circ m_1(F_{10})}{t}(y)dy \to \int_{\underline{y}}^{y_1} F'_{20}(F_{10}(y))h_1(y)dy.$$

The right-hand side is linear with respect to h_1. It is also continuous since the integral is taken over a bounded interval. The second point follows.

Lemma B.5 *Assume Assumptions 1-5, 7, 15-17 hold. Let* $\theta = (F_{000}, ..., F_{011}, F_{100}, ..., F_{111}, \lambda_0, \mu_0, \lambda_1, \mu_1)$. *For* $d \in \{0,1\}$ *and* $q \in \mathcal{Q}$, $\theta \mapsto \int_{\underline{y}}^{\overline{y}} \underline{B}_d(y)dy$, $\theta \mapsto \int_{\underline{y}}^{\overline{y}} \overline{B}_d(y)dy$, $\theta \mapsto \overline{B}_d^{-1}(q)$ *and* $\theta \mapsto \underline{B}_d^{-1}(q)$ *are Hadamard differentiable tangentially to* $(\mathcal{C}^0)^4 \times \mathbb{R}^2$.

Proof: the proof is complicated by the fact that even if the primitive cdf are smooth, the bounds $\underline{B}_d$ and $\overline{B}_d$ may admit kinks, so that Hadamard differentiability is not trivial to derive. The proof is also lengthy as $\underline{B}_d$ and $\overline{B}_d$ take different forms depending on $d \in \{0,1\}$ and whether $\lambda_0 < 1$ or $\lambda_0 > 1$. Before considering all possible cases, note that by Assumption 7, $\underline{B}_d = C_d(\underline{T}_d)$.

1. Lower bound $\underline{B}_d$

For $d \in \{0,1\}$, let $U_d = \frac{\lambda_d F_{d01} - H_d^{-1}(m_1(\mu_d F_{d11}))}{\lambda_d - 1}$, so that

$$\begin{aligned} \underline{T}_d &= M_0\left(m_1\left(U_d\right)\right), \\ C_d(\underline{T}_d) &= \frac{\mu_d F_{d11} - H_d\left(\lambda_d F_{d01} + (1-\lambda_d)\underline{T}_d\right)}{\mu_d - 1}. \end{aligned}$$

Also, let

$$y^u_{0d} = \inf\{y : U_d(y) > 0\} \text{ and } y^u_{1d} = \inf\{y : U_d(y) > 1\}.$$

When y^u_{0d} and y^u_{1d} are in $\mathbb{R}$, we have, by continuity of U_d, $U_d(y^u_{0d}) = 0$ and $U_d(y^u_{1d}) = 1$. Consequently, $\underline{T}_d(y^u_{0d}) = U_d(y^u_{0d})$ and $\underline{T}_d(y^u_{1d}) = U_d(y^u_{1d})$.

Case 1: $\lambda_0 < 1$ *and* $d = 0$.

In this case, $U_0 = \frac{H_0^{-1}(\mu_0 F_{011}) - \lambda_0 F_{001}}{1-\lambda_0}$. We first prove by contradiction that $y_{00}^u = +\infty$. First, because $\lim_{y\to+\infty} U_0(y) < 1$, we have

$$\lim_{y\to+\infty} \underline{T}_0(y) = M_0(\lim_{y\to+\infty} U_0(y)) < 1.$$

Thus, by Assumption 7, $U_0(y) < 1$ for all y, otherwise $\underline{T}_0(y)$ would be decreasing. Hence, $y_{10}^u = +\infty$.

Therefore, when $y_{00}^u < +\infty$, there exists y such that $0 < U_0(y) < 1$. Assume that there exists $y' \geq y$ such that $U_0(y') < 0$. By continuity and the intermediate value theorem, this would imply that there exists $y'' \in (y, y')$ such that $U_0(y'') = 0$. But since both $U_0(y)$ and $U_0(y'')$ are included in $[0, 1]$, this would imply that $\underline{T}_0$ is strictly decreasing between y and y'', which is not possible under Assumption 7. This proves that when $y_{00}^u < +\infty$, there exists y such that for every $y' \geq y$, $0 \leq U_0(y') < 1$.

Consequently, $\underline{T}_0 = U_0$ for every $y' \geq y$. This in turn implies that $C_0(\underline{T}_0) = 0$ for every $y' \geq y$. Moreover, $C_0(\underline{T}_0)$ is increasing under Assumption 7, which implies that $C_0(\underline{T}_0) = 0$ for every y. This proves that when $y_{00}^u < +\infty$, $C_0(\underline{T}_0) = 0$. This implies that $\mathcal{S}_0$ is empty, which violates Assumption 7. Therefore, under Assumption 7, we cannot have $y_{00}^u < +\infty$ when $\lambda_0 < 1$. Because $y_{00}^u = +\infty$, $\underline{T}_0 = 0$. Therefore,

$$C_0(\underline{T}_0)(y) = \frac{\mu_0 F_{011}(y) - H_0\left(\lambda_0 F_{001}(y)\right)}{\mu_0 - 1}.$$

The map $F \mapsto \int_{\mathcal{S}(Y)} F(y)dy$ is linear and continuous with respect to the supremum norm at any continuous F because $\mathcal{S}(Y)$ is bounded. It is thus Hadamard differentiable, tangentially to $\mathcal{C}^0$. Therefore, by Assumption 17, the first point of Lemma B.4, and the chain rule,

$$\theta \mapsto \int_{\mathcal{S}(Y)} \underline{B}_0(y)dy$$

is Hadamard differentiable tangentially to $(\mathcal{C}^0)^4 \times \mathbb{R}^2$.

Then, the map $F \mapsto F^{-1}$ is Hadamard differentiable at any F with strictly positive derivative, tangentially to $\mathcal{C}^0$ (see, e.g., van der Vaart, 2000, Lemma 21.4). Moreover, by Assumption 17, $C_0(\underline{T}_0)$ is increasing and differentiable with strictly positive derivative on $\underline{S}_0$, which is equal to $\mathcal{S}(Y)$ in this case. Thus, by the first point of Lemma B.4 and the chain rule, $\theta \mapsto \underline{B}_0^{-1}(q)$ is Hadamard differentiable tangentially to $(\mathcal{C}^0)^4 \times \mathbb{R}^2$ for any $q \in \mathcal{Q}$.

Case 2: $\lambda_0 > 1$ *and* $d = 0$.

In this case,

$$U_0 \quad = \quad \frac{\lambda_0 F_{001} - H_0^{-1}(\mu_0 F_{011})}{\lambda_0 - 1}.$$

Therefore, $\lim_{y \to \underline{y}} U_0(y) = 0$, and $\lim_{y \to \overline{y}} U_0(y) > 1$. As a result, $-\infty < y_{10}^u < +\infty$, and $\underline{T}_0(y_{10}^u) = U_0(y_{10}^u) = 1$. This in turn implies $C_0(\underline{T}_0)(y_{10}^u) = 0$. Combining this with Assumption 7 implies that $C_0(\underline{T}_0)(y) = 0$ for every $y \leq y_{10}^u$. Moreover, Assumption 7 also implies that $\underline{T}_d(y) = 1$ for every $y \geq y_{10}^u$. Therefore,

$$C_0(\underline{T}_0)(y) = \left| \begin{array}{ll} 0 & \text{if } y \leq y_{10}^u, \\ \frac{\mu_0 F_{011}(y) - H_0(\lambda_0 F_{001}(y) + (1-\lambda_0))}{\mu_0 - 1} & \text{if } y > y_{10}^u. \end{array} \right.$$

Thus, $C_0(\underline{T}_0)(y) = M_0(R_2(F_{011}, F_{010}, F_{000}, F_{001}, \lambda_0, \mu_0))$, where R_2 is defined as in Lemma B.4. Hadamard differentiability of $\int_{\underline{y}}^{\overline{y}} C_0(\underline{T}_0)(y)dy$ tangentially to $(\mathcal{C}^0)^4 \times \mathbb{R}^2$ thus follows by Points 1 and 2 of Lemma B.4, the chain rule and the fact that by Assumption 16, $(F_{011}, F_{010}, F_{000}, F_{001}, \lambda_0, \mu_0) \in (\mathcal{C}^1)^4 \times [0, \infty) \times ([0, \infty) \backslash \{1\})$. As for the QTE, note that by Point 1 of Lemma B.4, $\theta \mapsto C_0(\underline{T}_0)$ is Hadamard differentiable as a function on $(y_{10}^u, \overline{y})$, tangentially to $(\mathcal{C}^0)^4 \times \mathbb{R}^2$. By Assumption 17, $C_0(\underline{T}_0)$ is also strictly increasing and differentiable with positive derivative on $\underline{S}_0 = (y_{10}^u, \overline{y})$. Thus, by point 1 of Lemma B.4, Hadamard differentiability of $F \mapsto F^{-1}(q)$ at $(C_0(\underline{T}_0), q)$ for $q \in \mathcal{Q}$ tangentially to $\mathcal{C}^0$, and the chain rule, $\theta \mapsto \underline{B}_0^{-1}(q)$ is Hadamard differentiable tangentially to $(\mathcal{C}^0)^4 \times \mathbb{R}^2$.

Case 3: $\lambda_0 < 1$ *and* $d = 1$.

In this case,

$$U_1 \quad = \quad \frac{\lambda_1 F_{100} - H_1^{-1}(\mu_1 F_{111})}{\lambda_1 - 1}.$$

$\mu_1 > 1$ implies that $\frac{1}{\mu_1} < 1$. Therefore, $y^* = F_{111}^{-1}(\frac{1}{\mu_1})$ is in $\mathring{\mathcal{S}}(Y)$ under Assumption 4.

Case 3.a: $\lambda_0 < 1$, $d = 1$ *and* $y_{01}^u < y^*$.

We have $U_1(y^*) = \frac{\lambda_1 F_{100}(y^*) - 1}{\lambda_1 - 1} < 1$. Assume that $U_1(y^*) < 0$. Since $y_{01}^u < y^*$, this implies that there exists $y < y^*$ such that $0 < U_1(y)$. Since U_1 is continuous, there also exists $y' < y^*$ such that $0 < U_1(y') < 1$. By continuity and the intermediate value theorem, this finally implies that there exists y'' such that $y' < y''$ and $U_1(y'') = 0$. This contradicts Assumption 7 since this would imply that $\underline{T}_1$ is decreasing between y' and y''. This proves that

$$0 \leq U_1(y^*) < 1.$$

Therefore, $\underline{T}_1(y^*) = U_1(y^*)$, which in turn implies that $C_1(\underline{T}_1)(y^*) = 0$. By Assumption 7, this implies that for every $y \leq y^*$, $C_1(\underline{T}_1)(y) = 0$.

For every y greater than y^*,

$$U_1(y) \quad = \quad \frac{\lambda_1 F_{100}(y) - 1}{\lambda_1 - 1}.$$

$U_1(y) < 1$. Since $U_1(y^*) \geq 0$ and $y \mapsto \frac{\lambda_1 F_{100}(y)-1}{\lambda_1 - 1}$ is increasing, $U_1(y) \geq 0$. Consequently, for $y \geq y^*$, $\underline{T}_1(y) = U_1(y)$.

Finally, we obtain

$$C_1(\underline{T}_1)(y) = \left| \begin{array}{ll} 0 & \text{if } y \leq y^*, \\ \frac{\mu_1 F_{111}(y)-1}{\mu_1 - 1} & \text{if } y > y^*. \end{array} \right.$$

The result follows as in Case 2 above.

Case 3.b: $\lambda_0 < 1$, $d = 1$ *and* $y_{01}^u \geq y^*$.

For all $y \geq y^*$, $U_1(y) = \frac{\lambda_1 F_{100}(y)-1}{\lambda_1-1}$. This implies that $y_{01}^u = F_{100}^{-1}(1/\lambda_1) < +\infty$ and $U_1(y_{01}^u) = 0$. Because $y \mapsto \frac{\lambda_1 F_{100}(y)-1}{\lambda_1-1}$ is increasing, $U_1(y) \geq 0$ for every $y \geq y_{01}^u$. Moreover, $U_1(y) \leq 1$. Therefore, $\underline{T}_1(y) = U_1(y)$ for every $y \geq y_{01}^u$. Beside, for every y lower than y_{01}^u, $\underline{T}_1(y) = 0$. As a result,

$$C_1(\underline{T}_1)(y) = \left| \begin{array}{ll} \frac{\mu_1 F_{111}(y) - H_1(\lambda_1 F_{101}(y))}{\mu_1 - 1} & \text{if } y \leq y_{01}^u, \\ \frac{\mu_1 F_{111}(y)-1}{\mu_1-1} & \text{if } y > y_{01}^u. \end{array} \right.$$

This implies that

$$\int_{\underline{y}}^{\overline{y}} C_1(\underline{T}_1)(y)dy = \frac{1}{\mu_1 - 1}\left[\mu_1 \int_{\underline{y}}^{\overline{y}} F_{111}(y)dy - R_4(\lambda_1 F_{101}, H_1)\right],$$

where R_4 is defined in Lemma B.4. $\theta \mapsto \int_{\underline{y}}^{\overline{y}} F_{111}(y)dy$ is Hadamard differentiable at F_{111}, tangentially to $\mathcal{C}^0$. As shown in the proof of Lemma B.4, $H_1 = F_{110} \circ F_{100}^{-1}$ is a Hadamard differentiable function of (F_{110}, F_{100}), tangentially to $(\mathcal{C}^0)^2$. Thus, by Lemma B.4 and the chain rule, $R_4(\lambda_1 F_{101}, H_1)$ is a Hadamard differentiable function of $(F_{101}, F_{110}, F_{100})$, tangentially to $(\mathcal{C}^0)^3$. The result follows for $\int_{\underline{y}}^{\overline{y}} C_1(\underline{T}_1)(y)dy$.

The previous display also shows that $C_1(\underline{T}_1)$ is Hadamard differentiable as a function of $(F_{100}, F_{101}, F_{110}, F_{111}, \lambda_1, \mu_1)$ when considering the restriction of these functions to $(\underline{y}, y_{01}^u)$ only. By Assumption 17, $C_1(\underline{T}_1)$ is also a differentiable function with positive derivative on $(\underline{y}, y_{01}^u)$. Therefore, using once again the first point of Lemma B.4 and the chain rule, $\theta \mapsto C_1(\underline{T}_1)^{-1}(q)$ is Hadamard differentiable tangentially to $(\mathcal{C}^0)^4 \times \mathbb{R}^2$, for $q \in (C_1(\underline{T}_1)(\underline{y}), C_1(\underline{T}_1)(y_{01}^u)) = (0, q_1)$. The same holds when considering the interval $(y_{01}^u, \overline{y})$ instead of $(\underline{y}, y_{01}^u)$. Hence, $\theta \mapsto \underline{B}_1^{-1}(q)$ is Hadamard differentiable tangentially to $(\mathcal{C}^0)^4 \times \mathbb{R}^2$, for $q \in (0,1)\backslash\{q_1\} = \mathcal{Q}$.

Case 4: $\lambda_0 > 1$ *and* $d = 1$.

In this case,

$$U_1 = \frac{H_1^{-1}(\mu_1 F_{111}) - \lambda_1 F_{100}}{1-\lambda_1}.$$

Therefore, $\lim_{y \to \underline{y}} U_1(y) = 0$, which implies that $y_{11}^u > -\infty$. As above, $\mu_1 > 1$ implies that y^* is in $\overset{\circ}{\mathcal{S}}(Y)$ under Assumption 4. $U_1(y^*) = \frac{1-\lambda_1 F_{100}(y^*)}{1-\lambda_1} > 1$, which implies that $y_{11}^u < +\infty$. Therefore, reasoning as for Case 2, we obtain

$$C_1(\underline{T}_1)(y) = \begin{vmatrix} 0 & \text{if } y \leq y_{11}^u, \\ \frac{\mu_1 F_{111}(y) - H_1(\lambda_1 F_{100}(y) + (1-\lambda_1))}{\mu_1 - 1} & \text{if } y > y_{11}^u. \end{vmatrix}$$

The result follows as in Case 2 above.

2. Upper bound $\overline{B}_d$.

Let $V_d = \frac{\lambda_d F_{d01} - H_d^{-1}(M_0(\mu_d F_{d11} + (1-\mu_d)))}{\lambda_d - 1}$, so that

$$\begin{aligned} \overline{T}_d &= M_0(m_1(V_d)), \\ C_d(\overline{T}_d) &= \frac{\mu_d F_{d11} - H_d\left(\lambda_d F_{d01} + (1-\lambda_d)\overline{T}_d\right)}{\mu_d - 1}. \end{aligned}$$

Also, let

$$y_{0d}^v = \inf\{y : V_d(y) > 0\}, \quad y_{1d}^v = \inf\{y : V_d(y) > 1\}.$$

Note that when y_{0d}^v and y_{1d}^v are in $\mathbb{R}$, by continuity of V_d we have $V_d(y_{0d}^v) = 0$ and $V_d(y_{1d}^v) = 1$. Consequently, $\overline{T}_d(y_{0d}^v) = V_d(y_{0d}^v)$ and $\overline{T}_d(y_{1d}^v) = V_d(y_{1d}^v)$.

Case 1: $\lambda_0 < 1$ and $d = 0$.

In this case,

$$V_0 = \frac{H_0^{-1}(\mu_0 F_{011} + (1-\mu_0)) - \lambda_0 F_{001}}{1-\lambda_0}.$$

Since $\mu_0 < 1$, $\lim_{y \to \underline{y}} V_0(y) > 0$ and can even be greater than 1.

First, let us prove by contradiction that $y_{10}^v = -\infty$. $V_0(y) \leq 1$ for every $y \leq y_{10}^v$. Using the fact that $\lim_{y \to \underline{y}} V_0(y) > 0$ and that $\overline{T}_0$ must be increasing under Assumption 7, one can also show that $0 \leq V_0(y)$ for every $y \leq y_{10}^v$. This implies that $\overline{T}_0(y) = V_0(y)$ which in turn implies that $C_0(\overline{T}_0)(y) = 1$ for every $y \leq y_{10}^v$. Since $C_0(\overline{T}_0)$ must be increasing under Assumption 7, this implies that for every $y \in \mathcal{S}(Y)$,

$$C_0(\overline{T}_0)(y) = 1.$$

This implies that $\mathcal{S}_0$ is empty, which violates Assumption 7. Therefore, $y_{10}^v = -\infty$.

$y_{10}^v = -\infty$ implies that $\lim_{y \to \underline{y}} \overline{T}_0(y) = 1$. This combined with Assumption 7 implies that $\overline{T}_0(y) = 1$ for every $y \in \mathcal{S}(Y)$. Therefore,

$$C_0(\overline{T}_0)(y) = \frac{\mu_0 F_{011}(y) - H_0\left(\lambda_0 F_{001}(y) + (1-\lambda_0)\right)}{\mu_0 - 1}.$$

The result follows as in Case 1 of the lower bound.

Case 2: $\lambda_0 > 1$ and $d = 0$.

In this case,

$$V_0 = \frac{\lambda_0 F_{001} - H_0^{-1}(\mu_0 F_{011} + (1-\mu_0))}{\lambda_0 - 1}.$$

Since $\mu_0 < 1$, $\lim_{y \to \underline{y}} V_0(y) < 0$. Therefore, $y_{00}^v > -\infty$.

Case 2.a): $\lambda_0 > 1$, $d = 0$ and $y_{00}^v < +\infty$.

If $y_{00}^v \in \mathbb{R}$, $\overline{T}_0(y_{00}^v) = V_0(y_{00}^v)$ which in turn implies that $C_0(\overline{T}_0)(y_{00}^v) = 1$. By Assumption 7, this implies that for every $y \geq y_{00}^v$, $C_0(\overline{T}_0)(y) = 1$. For every $y \leq y_{00}^v$, $\overline{T}_0(y) = 0$, so that

$$C_0(\overline{T}_0) = \frac{\mu_0 F_{011} - H_0\left(\lambda_0 F_{001}\right)}{\mu_0 - 1}.$$

As a result,

$$C_0(\overline{T}_0)(y) = \left| \begin{array}{ll} \frac{\mu_0 F_{011}(y) - H_0(\lambda_0 F_{001}(y))}{\mu_0 - 1} & \text{if } y \leq y_{00}^v, \\ 1 & \text{if } y > y_{00}^v. \end{array} \right.$$

The result follows as in Case 2 of the lower bound.

Case 2.b): $\lambda_0 > 1$, $d = 0$ and $y_{00}^v = +\infty$.

If $y_{00}^v = +\infty$, $\overline{T}_0(y) = 0$ for every $y \in \mathcal{S}(Y)$, so that

$$C_0(\overline{T}_0)(y) = \frac{\mu_0 F_{011}(y) - H_0\left(\lambda_0 F_{001}(y)\right)}{\mu_0 - 1}.$$

The result follows as in Case 1 of the lower bound.

Case 3: $\lambda_0 < 1$ and $d = 1$.

In this case,

$$V_1 = \frac{\lambda_1 F_{101} - H_1^{-1}(\mu_1 F_{111} - (\mu_1 - 1))}{\lambda_1 - 1}.$$

Therefore, $\lim_{y \to \underline{y}} V_1(y) = 0$, which implies that $y_{11}^v > -\infty$. $\mu_1 > 1$ implies that $\frac{\mu_1 - 1}{\mu_1} < 1$. Therefore, $y^* = F_{111}^{-1}(\frac{\mu_1 - 1}{\mu_1})$ is in $\overset{\circ}{\mathcal{S}}(Y)$ under Assumption 4.

Case 3.a): $\lambda_0 < 1$, $d = 1$ and $y_{11}^v > y^*$.

We have $V_1(y^*) = \lambda_1 F_{101}(y^*)/(\lambda_1 - 1) > 0$. If $y^* < y^v_{11}$, $V_1(y^*) < 1$. Therefore, $0 < \overline{T}_1(y^*) = V_1(y^*) < 1$. This implies that $C_1(\overline{T}_1)(y^*) = 1$ which in turn implies that $C_1(\overline{T}_1)(y) = 1$ for every $y \geq y^*$ under Assumption 7.

For every y lower than y^*,

$$V_1(y) = \frac{\lambda_1 F_{101}(y)}{\lambda_1 - 1}.$$

$V_1(y) > 0$. Since by assumption $y^v_{11} > y^*$, $V_1(y) < 1$. Consequently, for $y \leq y^*$, we have $\overline{T}_1(y) = V_1(y)$. As a result,

$$C_1(\overline{T}_1)(y) = \left| \begin{array}{ll} \frac{\mu_1 F_{111}(y)}{\mu_1 - 1} & \text{if } y \leq y^*, \\ 1 & \text{if } y > y^*. \end{array} \right.$$

The result follows as in Case 2 of the lower bound.

Case 3.b): $\lambda_0 < 1$, $d = 1$, and $y^v_{11} \leq y^*$.

First, $V_1(y^v_{11}) = 1$, implying $\overline{T}_1(y^v_{11}) = 1$. By Assumption 7, $\overline{T}_1(y) = 1$ for all $y \geq y^v_{11}$. Second, if $y \leq y^v_{11} \leq y^*$, $V_1(y) = \frac{\lambda_1 F_{101}(y)}{\lambda_1 - 1}$. Thus V_1 is increasing on $(-\infty, y^v_{11})$. Moreover $V_1(y^v_{11}) = 1$. Hence, $V_1(y) \leq 1$ for every $y \leq y^v_{11}$. Because we also have $V_1(y) \geq 0$, $\overline{T}_1(y) = V_1(y)$ for every $y \leq y^v_{11}$.

As a result,

$$C_1(\overline{T}_1)(y) = \left| \begin{array}{ll} \frac{\mu_1 F_{111}(y)}{\mu_1 - 1} & \text{if } y \leq y^v_{11}, \\ \frac{\mu_1 F_{111}(y) - H_1(\lambda_1 F_{101}(y) + 1 - \lambda_1)}{\mu_1 - 1} & \text{if } y > y^v_{11}. \end{array} \right.$$

The result follows as in Case 3.b) of the lower bound. Note that here, $C_1(\overline{T}_1)(y)$ is kinked at y^v_{11}, with $C_1(\overline{T}_1)(y^v_{11}) = q_2$. Hence, we have to exclude this point of the domain on which $\theta \mapsto \overline{B}_1^{-1}(q)$ is Hadamard differentiable.

Case 4: $\lambda_0 > 1$ and $d = 1$.

In this case,

$$V_1 = \frac{H_1^{-1}(\mu_1 F_{111} - (\mu_1 - 1)) - \lambda_1 F_{101}}{1 - \lambda_1}.$$

$\lim_{y \to \overline{y}} V_1(y) = 1$, which implies that $y^v_{01} < +\infty$. As above, $\mu_1 > 1$ implies that $\frac{\mu_1 - 1}{\mu_1} < 1$. Therefore, $y^* = F^{-1}_{111}(\frac{\mu_1 - 1}{\mu_1})$ is in $\overset{\circ}{\mathcal{S}}(Y)$ under Assumption 4. $V_1(y^*) = \frac{-\lambda_1 F_{101}(y^*)}{1 - \lambda_1} < 0$. Since $\overline{T}_1$ is increasing under Assumption 7, one can show that this implies that $y^v_{01} > y^*$. Therefore, reasoning as for Case 2, we obtain that

$$C_1(\overline{T}_1)(y) = \left| \begin{array}{ll} \frac{\mu_1 F_{111}(y) - H_1(\lambda_1 F_{101}(y))}{\mu_1 - 1} & \text{if } y \leq y^v_{01}, \\ 1 & \text{if } y > y^v_{01}. \end{array} \right.$$

The result follows as in Case 2 of the lower bound.

C Effectiveness of a smoking cessation treatment

Smoking rate among the adult population in France is around 30%. This is much higher than in most western countries (see e.g. Beck et al., 2007). Varenicline, a pharmacotherapy for smoking cessation, has been marketed in France since February 2007. Randomized controlled trials (RCT) have shown Varenicline to be more efficient than other pharmacotherapies used in smoking cessation (see e.g. Jorenby et al., 2006). However, there have been few studies based on non experimental data to confirm the efficacy of this new drug in real life settings. Moreover, studies on this new drug only investigated its average effect, and none considered potentially heterogeneous effects.

In our analysis, we use a database from 17 French smoking cessation clinics, in which doctors, nurses, and psychologists help smokers quit. When a patient comes for the first time, the clinic staff asks her how many cigarettes she smokes per day in order to assess how serious her addiction is. They also measure the number of carbon monoxide (CO) parts per million in the air she expires using a CO meter. This latter measure is a biomarker for recent tobacco use. After collecting those measures and discussing with the patient, they may advise her treatments, such as nicotine replacement therapies, to help her quit. Patients then come back for follow-up visits. During those visits, CO measures are made to validate tobacco abstinence. This measure is much more reliable than daily cigarettes smoked, because it is not self-reported. Below 5 parts per million, a patient is regarded as a non smoker by clinics staff. She is regarded as a light smoker when her CO is between 5 and 10, as a smoker when it is between 10 and 20, and as a heavy smoker when it is above 20.[10] Therefore, a patient with a CO of 25 at follow-up not only failed to quit but is still a heavy smoker.

The rate of prescription of Varenicline is heterogeneous across clinics, as it ranges from 0% to 37%. A very strong predictor of clinics propensity to prescribe varenicline is the share of their staff holding the "diplome universitaire de tabacologie" in 2005-2006, i.e. before varenicline was released. The "diplome universitaire de tabacologie" is a university degree awarded to staff who followed a 60 hours course on how to help smokers quit. Doctors, nurses, and psychologists working in those clinics are not obliged to take this course, but a large fraction of them do. The share of staff holding this degree is also heterogeneous across clinics, as it ranges from 0 to 100%, with a median equal to 60%. The correlation between prescription rate and share of staff holding this degree is equal to 0.63. Staff who took this training a few years before varenicline got market approval might have then been told that preliminary RCT showed this new drug to be very promising. They might also have a stronger taste for learning

[10] Typically, light smokers smoke less than 10 cigarettes a day, smokers smoke between 10 and 20 cigarettes a day, and heavy smokers smoke more than 20 cigarettes.

than those who do not take this training, and be therefore more aware of medical innovations, and more enclined to adopting them.

We use the share of staff holding this degree in 2005-2006 to construct two groups of "control" and "treatment" clinics. Control clinics are those belonging to the first tercile of this measure, while treatment clinics are those belonging to the third tercile. Period 0 covers the 2 years before the release of Varenicline (February 2005 to January 2007), while period 1 extends over the 2 years following it (February 2007 to January 2009). Our sample is made up of the 7,468 patients who attended control and treatment clinics over these two periods and who came to at least one follow-up visit.

By construction, the prescription rate of Varenicline is 0% in control and treatment clinics at period 0. At period 1, it is equal to 4.9% in control clinics and 25.4% in treatment clinics. This differential evolution of treatment rates over time across those two groups of clinics is the source of variation we use to measure the effect of varenicline. Since $P(D_{10} = 0) = 1$, $F_{Y_{11}(1)|C}(y)$ is point identified by $F_{Y_{11}|D=1}$. On the other hand, we have by construction $P(D_{00} = 1) = 0$ and $P(D_{01} = 1) > 0$ since we observe treated individuals in the control group at period 1. Thus, we are in the partially identified case and we rely on Theorem 3.3 to estimate bounds for $F_{Y_{11}(0)|C}$.

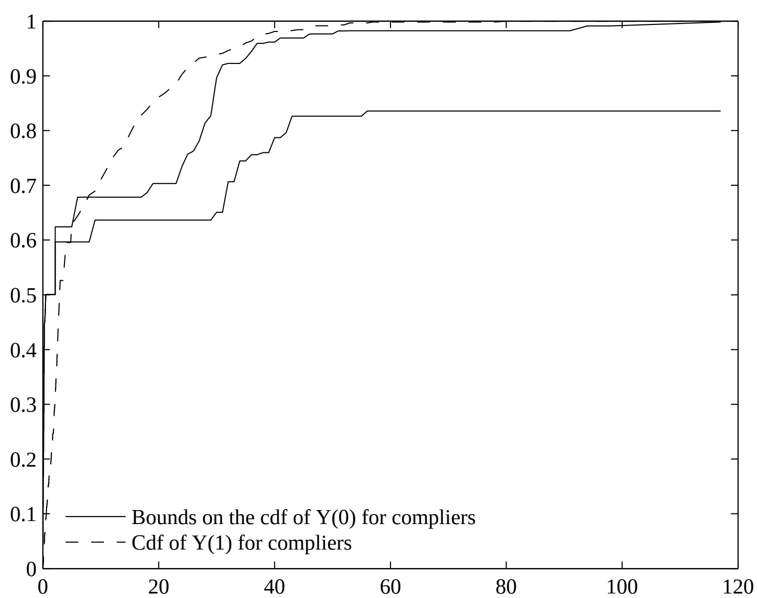

Figure 4: Estimated bounds on the cdf of $Y(0)$ and $Y(1)$ for compliers.

The resulting estimates are displayed in Figure 4. $\widehat{F}_{Y_{11}|D=1}$ dominates the estimate of the upper bound of $F_{Y_{11}(0)|C}(y)$ for all CO values above 10. This shows that varenicline reduces the share of high or very high CO values. This is reflected in Figure 5, which shows $\widehat{\underline{\tau}}_q$ and $\widehat{\overline{\overline{\tau}}}_q$, as well as the lower (resp. upper) bound of the 90% confidence interval of $\underline{\tau}_q$ (resp. $\overline{\tau}_q$). The two bounds are very close to 0 up to the 60th percentile, and both become strongly negative from that percentile onwards. We show bounds up to the 82% percentile only, because $\overline{q} = 0.82$.

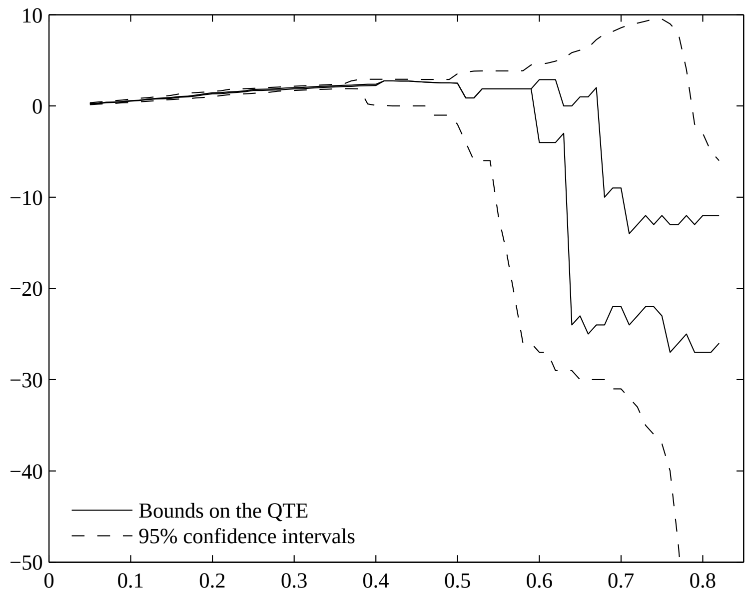

Figure 5: Estimated bounds for QTE on CO at follow-up.

In the previous figure, 0 most often lies in the confidence interval of τ_q, except for $q \in (0.78, 0.82)$. To increase statistical power, we include quartile of expired CO at baseline as a control variable. Results are displayed in Figure 6. τ_q is now significantly different from 0 for $q \in (0.62, 0.82)$.

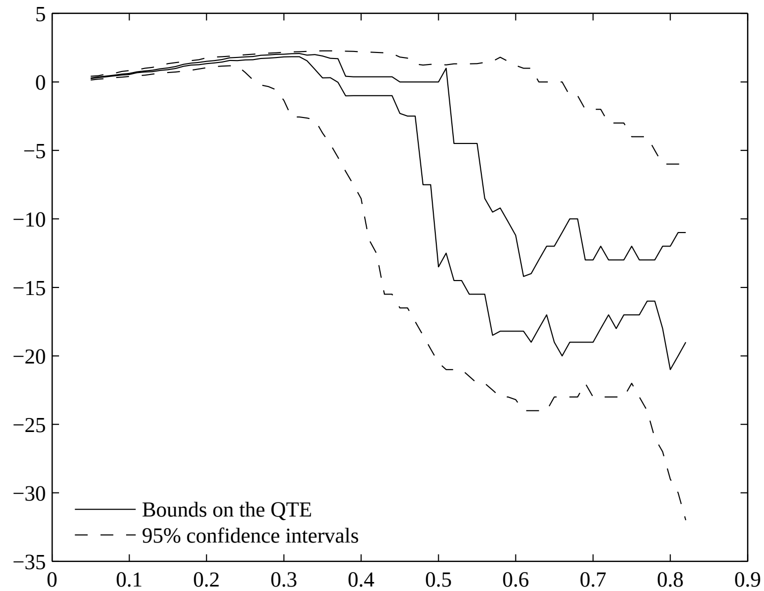

Figure 6: Estimated bounds for QTE, including baseline CO as a control.

Finally, we use our results to analyze the effect of varenicline on the probability of being a non smoker, a light smoker, a smoker, or an heavy smoker at follow-up. Results are displayed in Table 6. Varenicline does not significantly increase the share of non smokers at follow-up, even though our bounds point towards a small, positive effect. However, it has a large and significant negative effect on the share of heavy smokers: even as per our worst case upper bound, it decreases this share by 13.4 percentage points.

Table 6: Bounds on the effect of Varenicline on the shares of non smokers, light smokers, smokers, and heavy smokers

	$P(Y_{11}(1) \in I\|C) - P(Y_{11}(0) \in I\|C)$		
I	Lower bound	Upper bound	p-value
[0;5]	0.4%	19.5%	0.48
(5;10]	-19.1%	8.2%	0.93
(10;20]	4.5%	15.1%	0.27
(20;+∞)	-24.0%	-13.4%	0.02

Notes: the p-value corresponds to the test of the null hypothesis that $P(Y_{11}(1) \in I|C) - P(Y_{11}(0) \in I|C) = 0$.

The main assumption of our IV-CIC model states that the distribution of U_d, which can be interpreted here as addiction, or propensity to quit, is stable over time in the two groups. This could be violated, for instance if the most addicted patients come to treatment clinics in period 1 because they know they can get a prescription of Varenicline in those clinics and not in control ones. To assess the credibility of this assumption, we conduct the exact same analysis as in Figure 5, but with patients' baseline CO as our outcome. Baseline CO is indeed a good proxy for severity of addiction. The resulting bounds for those "placebo" QTE are displayed in Figure 7. $\widehat{\underline{\tau}}_q$ and $\widehat{\overline{\tau}}_q$ are small in absolute value; 0 is most often included between them, and it always lies within the confidence interval of τ_q. This supports the identifying assumption of our IV-CIC model.

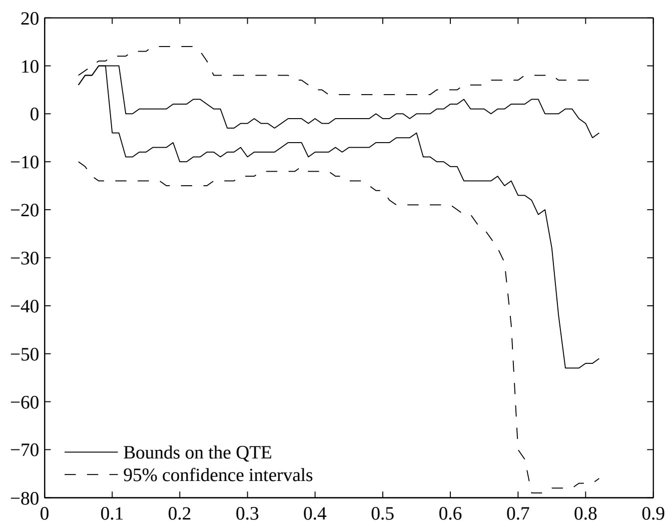

Figure 7: Bounds for QTE on baseline CO among compliers

References

Abadie, A. (2005), 'Semiparametric difference-in-differences estimators', *Review of Economic Studies* **72**(1), 1–19.

Akerman, A., Gaarder, I. & Mogstad, M. (2013), The skill complementarity of broadband internet, Technical report, Stockholm University, Department of Economics.

Altonji, J. & Blank, R. (2000), Race and gender in the labor market, *in* O. Ashenfelter & D. Card, eds, 'Handbook of Labor Economics', Amsterdam: Elsevier, pp. 3143–3259.

Andrews, D. W. K. & Soares, G. (2010), 'Inference for parameters defined by moment inequalities using generalized moment selection', *Econometrica* **78**(1), 119–157.

Angrist, J. D., Imbens, G. W. & Rubin, D. B. (1996), 'Identification of causal effects using instrumental variables', *Journal of the American Statistical Association* **91**(434), pp. 444–455.

Athey, S. & Imbens, G. W. (2002), Identification and inference in nonlinear difference-in-differences models, Technical report.

Athey, S. & Imbens, G. W. (2006), 'Identification and inference in nonlinear difference-in-differences models', *Econometrica* **74**(2), 431–497.

Beck, F., Guilbert, P. & Gautier, A. (2007), Baromètre Santé [French health barometer], Saint-Denis, INPES.

Blundell, R., Dias, M. C., Meghir, C. & Reenen, J. V. (2004), 'Evaluating the employment impact of a mandatory job search program', *Journal of the European Economic Association* **2**(4), 569–606.

Bonhomme, S. & Sauder, U. (2011), 'Recovering distributions in difference-in-differences models: A comparison of selective and comprehensive schooling', *Review of Economics and Statistics* **93**(2), 479–494.

Burgess, R. & Pande, R. (2005), 'Do rural banks matter? evidence from the indian social banking experiment', *American Economic Review* **95**(3), 780–795.

Chernozhukov, V., Lee, S. & Rosen, A. M. (2013), 'Intersection bounds: Estimation and inference', *Econometrica* **81**(2), 667–737.

de Chaisemartin, C. (2013), A note on the assumptions underlying instrumented difference in differences., Working paper.

D'Haultfoeuille, X., Hoderlein, S. & Sasaki, Y. (2013), Nonlinear difference-in-differences in repeated cross sections with continuous treatments. CEMMAP Working Paper CWP40/13.

Donald, S. G. & Lang, K. (2007), 'Inference with difference-in-differences and other panel data', *The Review of Economics and Statistics* **89**(2), 221–233.

Duflo, E. (2001), 'Schooling and labor market consequences of school construction in indonesia: Evidence from an unusual policy experiment', *American Economic Review* **91**(4), 795–813.

Field, E. (2007), 'Entitled to work: Urban property rights and labor supply in Peru', *The Quarterly Journal of Economics* **122**(4), 1561–1602.

Horowitz, J. L. & Manski, C. F. (1995), 'Identification and robustness with contaminated and corrupted data', *Econometrica* **63**(2), 281–302.

Imbens, G. W. & Angrist, J. D. (1994), 'Identification and estimation of local average treatment effects', *Econometrica* **62**(2), 467–75.

Imbens, G. W. & Manski, C. F. (2004), 'Confidence intervals for partially identified parameters', *Econometrica* **72**(6), 1845–1857.

Imbens, G. W. & Rubin, D. B. (1997), 'Estimating outcome distributions for compliers in instrumental variables models', *Review of Economic Studies* **64**(4), 555–574.

Jorenby, D. E., Hays, J. T., Rigotti, N. A., Azoulay, S. & Watsky, E. J. (2006), 'Efficacy of varenicline, an $\alpha 4\beta 2$ nicotinic acetylcholine receptor partial agonist, vs placebo or sustained-release bupropion for smoking cessation', *Journal of the American Medical Association* **306**(17).

Juhn, C., Murphy, K. M. & Pierce, B. (1993), 'Wage inequality and the rise in returns to skill', *Journal of Political Economy* **101**(3), 410–442.

Kitagawa, T. (2013), A bootstrap test for instrument validity in heterogeneous treatment effect models. CEMMAP Working Paper CWP53/13.

Lochner, L. & Moretti, E. (2004), 'The effect of education on crime: Evidence from prison inmates, arrests, and self-reports', *The American Economic Review* **94**(1), 155–189.

Manski, C. & Pepper, J. (2012), Partial identification of treatment response with data on repeated cross sections, Working papers.

Meyer, B. D., Viscusi, W. K. & Durbin, D. L. (1995), 'Workers' compensation and injury duration: Evidence from a natural experiment', *American Economic Review* **85**(3), 322–40.

Rosenbaum, P. R. & Rubin, D. B. (1983), 'The central role of the propensity score in observational studies for causal effects', *Biometrika* **70**(1), pp. 41–55.

Stoye, J. (2009), 'More on confidence intervals for partially identified parameters', *Econometrica* **77**(4), 1299–1315.

van der Vaart, A. W. (2000), *Asymptotic Statistics*, Cambridge Series in Statistical and Probabilistic Mathematics.

van der Vaart, A. W. & Wellner, J. A. (1996), *Weak convergence and Empirical Processes*, Springer.

Vytlacil, E. (2002), 'Independence, monotonicity, and latent index models: An equivalence result', *Econometrica* **70**(1), 331–341.